

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 2005, 04:50:04 ; Search time 1451 Seconds
(without alignments)
2304.211 Million cell updates/sec

Title: US-10-780-250-4

Perfect score: 69
Sequence: I gatgctaatactaggagt.....tactggtaatacactaca 69

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	46.6	67.5	5745	1	CPTOXIAB
2	37	53.6	4465	1	CSTOXING
3	33.8	49.0	4075	1	AE271719
4	33.8	49.0	4277	1	CLOCDTAB
5	31.2	45.2	807	6	AR395216
6	29.6	42.9	57660	1	AE016831
7	28	40.6	224280	2	AC123142
8	27.8	40.3	651	11	BV162185
9	27.6	40.0	43551	9	AL139012
10	27.6	40.0	116341	9	AL135936
11	27.6	40.0	180298	2	AL357633
12	27.6	40.0	211643	5	BX530036
13	27.6	40.0	225538	2	AC095519
14	27.6	40.0	238937	2	AC106652
15	27.4	39.7	227763	2	AC105534
16	27.2	39.4	65489	2	AC104345
17	27.2	39.4	143773	9	AC005631
18	27.2	39.4	163976	9	AC109630
19	27.2	39.4	193870	10	AL731674

20	27.2	39.4	194150	2	AC148825
21	27	39.1	90347	5	AL714030
22	27	39.1	99681	5	AL713850
23	27	39.1	160904	2	BX571969
24	27	39.1	173580	9	AL445928
25	27	39.1	226416	10	AC092752
26	26.8	38.8	5302	9	HS807111
27	26.8	38.8	112721	8	AP006401
28	26.8	38.8	127917	6	CQ861449
29	26.8	38.8	127917	9	HS257A7
30	26.8	38.8	146743	2	AC093588
31	26.8	38.8	153028	10	AL928739
32	26.8	38.8	155295	2	AC055859
33	26.8	38.8	166679	9	AC079899
34	26.8	38.8	172090	2	AC051661
35	26.8	38.8	172798	9	AC026585
36	26.8	38.8	176426	9	AC007370
37	26.8	38.8	183992	2	AC137252
38	26.8	38.8	221532	2	AC120251
39	26.8	38.8	235737	2	AC119689
40	26.8	38.8	248546	2	AC133690
41	26.8	38.8	292568	2	AC128486
42	26.6	38.6	121958	2	CR759861
43	26.6	38.6	175313	2	BX005160
44	26.6	38.6	201414	9	AC090644
45	26.6	38.6	219004	2	CR759887

ALIGNMENTS

RESULT 1

CPTOXIAB

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

REMARK

REFERENCE

AUTHORS

TITLE

JOURNAL

REMARK

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

SOURCE

1. 5745

/organism="Clostridium perfringens"

/mol_type="Genomic DNA"

/strain="NCIB 10748"

/sub_species="type E"

/db_xref="taxon:1502"

1453..1458

1465..2829

/codon_start=1

/transl_table=11

/product="iota toxin component Ia"

CPTOXIAB 5745 bp DNA linear BCT 09-SEP-2004
C.perfringens DNA for iota toxin polypeptides Ia and Ib.
X73562 GI:929031
iota toxin; iota toxin Ia; iota toxin Ib.
Clostridium perfringens
Clostridium perfringens
Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
Clostridium.

1 Perelle,S., Gilbert,M., Boquet,P. and Popoff,M.R.
Characterization of Clostridium perfringens iota-toxin genes and
expression in Escherichia coli

Infect. Immun. 61 (12), 5147-5156 (1993)
94041637
8225592

2 Popoff,M.R.
Direct Submission

Submitted (30-APR-1993) M.R. Popoff, Institut Pasteur, 28 rue du
Dr. Roux, Toxines Microbiennes, 75015 Paris Cedex 15, FRANCE

3 (bases 1 to 5745)
Popoff,M.R.

Direct Submission

Submitted (03-AUG-1995) M.R. Popoff, Institut Pasteur, 28 rue du
Dr. Roux, Toxines Microbiennes, 75015 Paris Cedex 15, FRANCE

On Aug 5, 1995 this sequence version replaced gi:414653.
Location/Qualifiers

1. 5745

/organism="Clostridium perfringens"

/mol_type="Genomic DNA"

/strain="NCIB 10748"

/sub_species="type E"

/db_xref="taxon:1502"

1453..1458

1465..2829

/codon_start=1

/transl_table=11

/product="iota toxin component Ia"

Matches	49;	Conservative	0;	Mismatches	20;	Indels	0;	Gaps	0;
Qy	1	GATGCTAATACTGTAGAGTTTCAATTTCACAGGGTGATCAGAAACGCCCTTTACTGGTAAT	60						
Db	2790	GATCAAAATACAGCTGAGTAGCAATTAATATTGCATATCAAATGGAATTTACTGGCAGT	2849						
Qy	61	ATCACTACA	69						
Db	2850	ATAACTACA	2858						
RESULT 3									
AF271719									
LOCUS									
DEFINITION	AF271719	Clostridium difficile CdtA (cdtA) and CdtB (cdtB) genes, complete cds.	4075 bp	DNA	linear	BCT 22-FEB-2002			
ACCESSION	AF271719	GI:8926246							
VERSION									
KEYWORDS		Clostridium difficile							
SOURCE		Clostridium difficile							
ORGANISM		Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae; Clostridium.							
REFERENCE		1. (bases 1 to 4075)							
AUTHORS		Chang,S.Y. and Song,K.P.							
TITLE		ADP-ribosylating binary toxin genes of Clostridium difficile strain CCUG 20309							
JOURNAL		DNA Seq. 12 (2), 115-120 (2001)							
MEDLINE		21604263							
PUBMED		11761709							
REFERENCE		2. (bases 1 to 4075)							
AUTHORS		Song,K.P. and Chang,S.Y.							
TITLE		Direct Submission							
JOURNAL		Submitted (24-MAY-2000) Microbiology, National University of Singapore, 5 Science Drive 2 117597, Singapore							
FEATURES									
source		1. .4075							
		/organism="Clostridium difficile"							
		/mol_type="genomic DNA"							
		/strain="CCUG 20309"							
		/db_xref="taxon:1496"							
gene		1. .1392							
		/gene="cdtA"							
CDS		1. .1392							
		/gene="cdtA"							
		/note="enzymatic domain; ADP-ribosylating binary toxin protein"							
		/codon_start=1							
		/transl_table=11							
		/product="CdtA"							
		/protein_id="AAF81760.1"							
		/db_xref="GI:8926247"							
		/translation="MKFRKHKISNCISILLIYLTGLGLLNNIYAODLOSYSKY CNTYYAPIERPSDFLDKKAEWKEARIEQLKERSEKALESYKDSVEISKY SONTFYFDYQIEANEREKEYELRNASKNDKPMVYVPFSPKFAFNKVIRTE QNRISEKFNEPKETIKFKDQGFKDLSYPGKGDEKPTPLMLHLPLRNTGMPL YTNNNVSTLI EOGYSIKIDKIVRIDGHYIKAAASVVSLDFDDVSKGSWGKA NYNWSNKLTLPNELADVDMRGYSTAINNYLISNGPVANNPELDSKITINEALKR EPIPTLVTVRRSGPOEFGTLTSPEYDFNKLKNIDAPKSWEGQALSYPNFISTSIG SVMSAPAKRKIVLRITIPKPGSGAYLSAIPGVAGEYVELLNHGSKFKINKIDSXDKG TITKLIVDATLIP"							
		1445. .4075							
gene		/gene="cdtB"							
CDS		1445. .4075							
		/gene="cdtB"							
		/note="binding domain; ADP-ribosylating binary toxin protein"							
		/codon_start=1							
		/transl_table=11							
		/product="CdtB"							
		/protein_id="AAF81761.1"							
		/db_xref="GI:8926248"							
		/translations="MKIQMNKKVL SFLTLTAIVSQALVVPVYAOSTSNHSHKKKEI"							

VNEDILPNNGLMGYFTDHFHDKDLKLMDIPKGNLKFEBKKYVDKLLDDKDSVKSIR
TGRIPSKOGEYTSLTDRODVLMQVNTESTISNTLKVMNMKGKEYKVRLELDQNGLG
IDNLSSPNIYWELDGMMKIIPBENLFRLDYSNIEKDPPFIPNNFPFKMLSMDEWED
LOTNDNIIPOSYERNGYTIKDLIAVKWEDSFAEQGYKYVSNYLESNTAGDPYTDYEK
ASGSFDKAICTEARDPLVAAYPIVGVTGMCKLIISTNEHASTDQGTQTVSRATNTSKTES
NTAGVSVNGYQNGFTANTYNVSHPTDNSTAVQDSGESWNLTGLSINKGESAYINAN
VRYANGTGAPMYKPTTNLVLDGDTLSTIKAEQNGICNNLSPDGYTPPKGLSPALAN
TMDFSRRIPIPYNDQLKLDAGKQIKLETQVSGFNFGTKNSGGIYPTGENSWSDXIIS
NDIPSASILIPTENESYERRVTAKNLQDPEDKTPELTIGEAIEKAFGATKQDGLLYF
QIDPIDESCVELIFDDNTANKIKDSLKTLSDKKIKIVNVLERGMNILLIKTPTYPTNFDD
YNNYPSTNVNTNKDGLQSGANKLGTETKIKIPMSSELKPVKRYVFVSGSKDPLTS
SIIVKIKEEKTDYLVPQGYTKFSYEFETTEKOSNNIEITLIGSETTYLDNLSITE
LNSTPILDEPEVKIPTDGEIIDAHKIYFADLNPFPSTGNTYINGMYPAPTQTNKEAL
DIYQKRVREDTLOYSGFQDGTOKDMERNYLGDPNQPKTNYNVLRSYFTGGENIMITYK
KLRIYAITPDRRELLIVSD"

Query Match 49.0%; Score 33.8; DB 1; Length 4075;
Best Local Similarity 68.1%; Pred. No. 1;
Matches 47; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

Qy 1 GATGCTAAACTCTAGGAGTTTCAAATTTCAGCAGCGGTATCAGAACGGCTTTACTCGTAAT 60
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2441 GAATCTAATACAGCTGGTGTGCTGTTAATGATAGGATATCAAAATGATTCACAGCTAAT 2500
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 61 ATCACTACA 69
|||||

Db 2501 GTAACTACA 2509
|||||

RESULT 4
CLOC DTAB 4277 bp DNA linear BCT 21-AUG-1997
LOCUS Clostridium difficile ADP-ribosyltransferase enzymatic and binding
DEFINITION component (cdta and cdtB) genes, complete cds's.
ACCESSION L76081
VERSION L76081.1 GI:2340080
KEYWORDS ADP-ribosyltransferase.
SOURCE Clostridium difficile
ORGANISM Clostridium difficile
Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
Clostridium.

REFERENCE 1 (bases 1 to 4277)
AUTHORS Perelle,S., Gibert,M., Bourlioux,P., Corthier,G. and Popoff,M.R.
TITLE Production of a complete binary toxin (actin-specific
ADP-ribosyltransferase) by Clostridium difficile CD196
JOURNAL Infect. Immun. 65 (4), 1402-1407 (1997)
MEDLINE 97230316
PUBMED 9119480
COMMENT GSDB:S:73984.

FEATURES
source Original source text: Clostridium difficile (strain CD196) DNA.
Location/Qualifiers
1..4277
/organism="Clostridium difficile"
/mol_type="genomic DNA"
/strain="CD196"
/db_xref="taxon:1496"
91..1487
/genes="cdta"
91..97
/genes="cdta"
/notes="putative"
105..1487
/genes="cdta"
/notes="enzymatic component; putative"
/codon_start=1
/transl_table=11
/product="ADP-ribosyltransferase"
/protein_id="AA067304.1"
/db_xref="GI:2340081"
/translation="MKKFRHKRISNCISILILYLTGLGLLPNNIYAODLOSYSSEKV
RNYYFYQIQAESREKEYKELRNAISKIDPMVYVYEPSEKPAFNKIVRTENQNE
CNTYKAPTEFLKDKKAKEWERKEARIEQKLRSEKALEYKOSVEISKYSQT
CNVYDQIQAESREKEYKELRNAISKIDPMVYVYEPSEKPAFNKIVRTENQNE

gene	ISLEKFNFEKTIQNLKFKODGFKDISLYBPGKGDEKPTPLMLHLKLPRNTGMLPYTN									
	TNNVSTLEQOYSIKDKI VPIVIDGKHVIKAEASVNSLD FKDDVSKGSGKANYN									
	DWNKLPNEADVNDVNRGGYTALNNVILISGPNPNPNEBELSKIINENALKREPI									
	PTNUTVVRSGPQFGLTSTSPFYDFNKLINIDAFKSKWEGQALSYPNFISTISGSVN									
RBS	MSAFKAKIVIRITIPKSPGAYLSAIPGVAGEYVLLNHGSKFKINKIDSYKDGTTT									
	KLIVDATLIP"									
	1528. .4170									
	/gene="cdtB"									
CDS	1528. .1532									
	/gene="cdtB"									
	/note="putative"									
	1540. .4170									
gene	/gene="cdtB"									
	/note="binding component; putative"									
	/codon_start=1									
	/transl_table=11									
RBS	/product="ADP-ribosyltransferase"									
	/protein_id="AAB67305.1"									
	/db_xref="GI:2340082"									
	/translation="MKIQMRNKKVLSFLTLTAIVSQALVYPVYAQTSTSNHSHKKKEL									
CDS	VNEDLNNGLMGYFSDHFKDKLMAPIKDGNLKFEEKVDKLLDKSDVKSIRW									
	TGRIPSKDGEYTLSDRDDVLMQVNTESITSLTKVMKKGEYKRIELQDKLGS									
	IDNLSSPNLYWELDMKKIIPEENLFLRDYSNIEKDDPIPNNNFFDPKLMSDWED									
	LDTDNNDIPDSYERNGYTIKDLIAVKWEDSPAEOGYKYVSNYLESNTAGDPYDYEK									
gene	ASSGFDKAIKTEARDPLVAAPYIVGVGMEKLIISTNEHASTDQKTVSRATTNSKTES									
	NTAGSVUNVGYQNGFTANVTNYSHTTDSNSTAVQDSGESWNTGLSNKGESAYINAN									
	VRYTNGAPMKIPTPTNLVDGDTLSTIKAQENQIGNNLSPGDTTPKGLSLPIALN									
	TYMDFSSLLPIINDQLKDGAKQIKLETTQVSGNFGTKNSGOIVTEGNSWSDYIS									
RBS	QIDISASIIIDLTENESYERRVTAKNLQDPEDKTPELTIGAEIKAFGATKKDGLLYF									
	NDIPIDESCVLIFDNDTANKIKDSLKTLSDKIIYNVKLERGMNLIKPTPTFFNFD									
	YNNPSTWSNVTNQDGLGSANKNGETKIKIPMSELKPYKRVFSGYKSDPLTIN									
	SLIVKIKAKEKTDYLVPEOGYKFSYEFETEKDSSNIEITLIGSTTVYLDNLSITE									
CDS	LNSPTEILDEVEKIPDQEIIMDAHKIYFADLNFPNSTGNTYINGMFAPTQTKNEAL									
	DYIQKRYEATLQYSGFKDICTGKDKEMNLYLGDENQPKNTVYNLRSYFTGGENIMTYK									
	KLRIYAITPDDRELLVLSVD"									
	ORIGIN									
Query Match	49.0%; Score 33.8; DB 1; Length 4277;									
	Best Local Similarity 68.1%; Pred. No. 1;									
	Matches 47; Conservative 0; Mismatches 22; Indels 0; Gaps 0;									
	ORIGIN									
Qy	1 GATGCTAATCTGTAGGAGTTTCAATTTACAGCAGGTATCAGAACGGCTTTACTGGTAAT 60									
	Db 2536 GAATCTAATACAGCTGGTGTCTCTTAAATGATAGATATCAAAATGGATTACAGCTAAT 2595									
Qy	61 ATCACTACA 69									
	Db 2596 GTAACCTACA 2604									
RESULT 5	AR395216/c									
	LOCUS									
	DEFINITION									
	Sequence 1231 from patent US 6617156. linear PAT 18-DEC-2003									
RBS	AR395216									
	ACCESSION									
	AR395216.1 GI:40123610									
	VERSION									
CDS	KEYWORDS									
	Unknown.									
	SOURCE									
	Unknown.									
gene	ORIGIN									
	Unclassified.									
	1 (bases 1 to 807)									
	Doucette-Stamm, L.A. and Bush, D.									
RBS	Nucleic acid and amino acid sequences relating to Enterococcus									
	faecalis for diagnostics and therapeutics									
	Patent: US 6617156-A 1231 09-SEP-2003;									
	Location/Qualifiers									
CDS	1. -807									
	source									
	/organism="unknown"									
	/mol_type="genomic DNA"									
ORIGIN	ORIGIN									
	Query Match 45.2%; Score 31.2; DB 6; Length 807;									
	Best Local Similarity 66.2%; Pred. No. 9.1;									
	ORIGIN									

Matches	45; Conservative	0; Mismatches	23; Indels	0; Gaps	0;
Qy	2	ATGCTAATCTGTAGGAGTTTCAATTTACAGCAGGTATCAGAACGGCTTTACTGGTAATA	61		
Db	475	ATGCGGCTTCACAGGAATAATAATTTTCATCAGACAGTGAACGCATTTACTGTTAATA	416		
Qy	62	TCACTACA	69		
Db	415	TACCTAAA	408		
LOCUS	AE016831 57660 bp DNA circular BCT 28-MAR-2003				
DEFINITION	Enterococcus faecalis V583 plasmid pTEF2, complete sequence.				
ACCESSION	AE016831				
VERSION	AE016831.1 GI:29345347				
KEYWORDS	.				
SOURCE	Enterococcus faecalis V583				
ORGANISM	Enterococcus faecalis V583				
REFERENCE	Bacteria; Firmicutes; Lactobacillales; Enterococcaceae;				
AUTHORS	Enterococcus				
	1 (bases 1 to 57660)				
	Paulsen, I., Banerjee, L., Myers, G., Nelson, K., Seshadri, R., Read, T.,				
	Fouts, D., Eisen, J., Gill, S., Heidelberg, J., Tettelin, H., Dodson, R.,				
	Unayam, L., Brinkac, L., Beanan, M., Daugherty, S., DeBoy, R.,				
	Durkin, A., Kolonay, J., Madupu, R., Nelson, W., Vamathevan, J.,				
	Tran, B., Upton, J., Hansen, T., Shetty, J., Khouri, H., Utterback, T.,				
	Radune, D., Ketchum, K., Dougherty, B. and Fraser, C.				
TITLE	Role of Mobile DNA in the Evolution of Vancomycin-Resistant				
JOURNAL	Enterococcus faecalis				
PUBMED	Science 299 (5615), 2071-2074 (2003)				
REFERENCE	12663927				
AUTHORS	2 (bases 1 to 57660)				
	Paulsen, I., Banerjee, L., Myers, G., Nelson, K., Seshadri, R., Read, T.,				
	Fouts, D., Eisen, J., Gill, S., Heidelberg, J., Tettelin, H., Dodson, R.,				
	Unayam, L., Brinkac, L., Beanan, M., Daugherty, S., DeBoy, R.,				
	Durkin, A., Kolonay, J., Madupu, R., Nelson, W., Vamathevan, J.,				
	Tran, B., Upton, J., Hansen, T., Shetty, J., Khouri, H., Utterback, T.,				
	Radune, D., Ketchum, K., Dougherty, B. and Fraser, C.				
TITLE	Direct Submission				
JOURNAL	Submitted (03-FEB-2003) The Institute for Genomic Research, 9712				
FEATURES	Medical Center Dr, Rockville, MD 20850, USA				
source	Location/Qualifiers				
	1..57660				
	/organism="Enterococcus faecalis V583"				
	/mol_type="genomic DNA"				
	/db_xref="taxon:226185"				
	/plasmid="pTEF2"				
gene	100..1110				
	/gene="repA-2"				
CDS	/locus_tag="EFB0001"				
	100..1110				
	/gene="repA-2"				
	/locus_tag="EFB0001"				
	/note="similar to GP.1041114; identified by sequence				
	similarity; putative"				
	/codon_start=1				
	/transl_table=11				
	/product="replication-associated protein RepA"				
	/protein_id="AAO83093.1"				
	/db_xref="GI:29345348"				
	/translation="WNDFQFISANETVRLNYQLPKVLFTSEFYKTMNSDIAYAML				
	QDRCEYSIQNWIDODGHYIFFTFRLMDILGCKENIKAKIKELKDKNLYEKRI				
	PKKSTGFEFKYPNKLYGLKEVSATDVAISNATYTFESGKNQPSDKDKIGVS				
	FESGKNQPSNNSTLTFESGKNQPNLYTNSLTDITDFEKRLOQOOLLQDQSE				
	VQETFTLSKLTIAAFSDTQIEAHMVGTIRAKTVKEVYNILVIGEDYQEEID				
	CLRWVHKIKTDSITVKSPKGLFYKSYFNLFVECALEKINKNNKSGSEITLHNWVE				
	NK"				
gene	1280..2437				
	/gene="trab-2"				
CDS	/locus_tag="EFB0003"				
	1280..2437				

/gene="trab-2"
/locus_tag="EFB0003"
/note="similar to GB:X06234, GB:M21005, GB:A12027,
SP:P05109, PID:34773, PID:386959, PID:818950, GB:X59842,
GB:D28769, GB:X80700, SP:P40425, PID:1841549, PID:35313,
PID:561658, and PID:634053; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="pheromone shutdown protein Trab"
/protein_id="AA083094.1"
/db_xref="GI:29345349"
/translation="MEEDIVLFYKGEIILGTSHTSAESADLVKTIQENPDPTIC
IEWQKRYKNIHDEWDDTDIVKIIKKNQPPVIFGVYKLFQKQVSDMNSLVGKE
FVAVDSKNIKIFYLIDRSDSLTFKRAMLNFREKVKLPYAFGKI FEGABETEE
VQNLSNPFPEELKESYPNLEWFEVTERDDYLATQNTANGKTVAVLGRAHLK
VQDLRLKNNQSDIQLKEIIIPKFKGSKLLEWIIIPGILLILGVSFGQTQVGEQLL
RWLLWNGLAALFTVFAIGHLTVLTSIIAPLATLPMVSVGVFSALVEATVRPKV
KDFQTMDDLQAKYIKTIYKNRVRVFLVFLSSLGGLGNIIGGLGIVKNLF"
2486..54075
/gene="trac-2"
/locus_tag="EFB0004"
2486..54075
/gene="trac-2"
/locus_tag="EFB0004"
/note="similar to GB:D10570, GB:M83215, GB:S60998,
SP:Q01196, PID:1932820, PID:530135, PID:557639,
PID:608133, PID:966995, PID:966997, PID:966999, GB:D10570,
GB:M83215, GB:S60998, SP:Q01196, PID:1932820, PID:530135,
PID:557639, PID:608133, PID:966995, PID:966997, and
PID:966999; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="Trac protein"
/protein_id="AA083095.1"
/db_xref="GI:29345350"
/translation="MLLVFASCGVKNKSTEDSKNTNKKVEQIATLSAGTPVQSLDPAT
AVDQTSVTLANVMEGLRYLDERKQPPQPAIAAGQPKISNGKTYTIVIRGAKWADGT
DITADDFVTAQVLDKPTKSPNVLELFAAIKNAKEISIGKQKETLGVKSKGKNTIEI
ELEPTPYTDLALTAVFPVQNAVKEYGEGYTKENIVTNGAFTLTDLNGVIGSD
KWTAKNPKYDKHVAKEIKFQVVDINTGILYNDGOLDPAVAGEYSKOLENNK
DFIRELSATWFEVNGRNKKSITSNKHARQINFAIDREAISNKILINGSIPAGVV
PSKLYNPKTKDFTNSLFDKSKAKDSWEKKBELGTDLSIDIMVNEEDLSKFL
GEYLNQDLDTGLKYSVTPATLQTERLNSGNFMIALSGWQADPADPVSFANFE
SKSLNHGCVANESYDILKNNSSKRLQELDAEKLILEDAGVIPLELQIGNAKLRNOK
ISEMKVHSIGAKYDYKWEIK"
complement(4101..5042)
/locus_tag="EFB0005"
complement(4101..5042)
/locus_tag="EFB0005"
/note="similar to GB:X15414, GB:J04795, GB:M34720,
SP:P15121, PID:178485, PID:178487, PID:178489, PID:178491,
PID:179036, PID:179038, PID:2228538, and PID:28647;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="transcriptional regulator, Cro/Ci family"
/protein_id="AA083096.1"
/db_xref="GI:29345351"
/translation="MDVKELLEVRKDLNFKKVDLYQEKSNKTYRIEILGEKSPTEI
ELISFTDTGLRSEFLYRGDITFLAASYGKRKIEVASLINNFELEKQFPPEIYKDR
FKNQCXYTLFLICIIISKLNSLYTFFKNQDIRELKNFYKRNIVILGADYAILSHLYM
APVYSNELDPTKLEKLIPIQOSYGLDIDHVAQLCKNAVTTFLEKDFEKADYLEQF
DIVKVSFLIDGTINLEMIYLKLRNPLKRNMEEYLKAVNIIVNLSQLGENDTYKE
LVKETSFIARKEFVSSLLNIADTFEGVGYLSDK"
5255..5332
/locus_tag="EFB0005.1"
5255..5332
/locus_tag="EFB0005.1"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"

/protein_id="AA083097.1"
/db_xref="GI:29345352"
/translation="MIKKSILKIKYVYVVLISLTILSLA"
5822..6202
/locus_tag="EFB0007"
5822..6202
/locus_tag="EFB0007"
/note="similar to GB:D38524, SP:P49902, PID:633071,
GB:D38524, SP:P49902, and PID:633071; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="probable pheromone-responsive regulatory protein
R"
/protein_id="AA083098.1"
/db_xref="GI:29345353"
/translation="MIELKATDHRITCOLTLENRQRFKLYNGTTERIVVDFSGRNLL
INPVSTEQDMERFFRQVKLIYFDGKVGVRGCGSELPLKLECFKMKVQRQKMLLN
APFNKVFEESENFREWCCKMRSYK"
6776..6964
/locus_tag="EFB0008"
6776..6964
/locus_tag="EFB0008"
/note="100% identity to the hypothetical protein prgt on
the Enterococcus faecalis plasmid pCF10 (PMID:1938961).;
similar to GB:M90360, and PID:184435; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA083099.1"
/db_xref="GI:29345354"
/translation="MTKKEQSIWRKEMLALMNEADADWYRNEDTERFKRIQELAKKIET
ASTROPSSHISKERFEAY"
6968..7081
/locus_tag="EFB0009"
6968..7081
/locus_tag="EFB0009"
/note="similar to GB:M90360, and PID:184435; identified
by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AA083100.1"
/db_xref="GI:29345355"
/translation="MGLQPKIEAEFHTTTTALQQWHKONGYPIYNKNRK"
7092..9767
/gene="prga"
/locus_tag="EFB0010"
7092..9767
/gene="prga"
/locus_tag="EFB0010"
/note="similar to GB:Z11692, GB:M19997, GB:X51466,
SP:P13639, PID:181969, PID:31106, GB:Z11692,
GB:M19997, GB:X51466, SP:P13639, PID:181969, PID:31106,
and PID:31108; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="surface exclusion protein Prga"
/protein_id="AA083101.1"
/db_xref="GI:29345356"
/translation="MKKIAGTGLSILVATGVAGIGNEVAAEQAPKTPENSSTEQP
TVKATQTTEQAITKQQVTEKQAIVDQKQVADTAKKEDADQSVKDDQAVVDQK
DALDQSQAVTDQQAQVVDKAKVDEATPSAIEKAKEQVATDTCQVDEQKQVDAQT
DVNQQAQVDEKAKETNAKQKQQAQVTAQKAEQAKLEELANAEKVAEKAEQ
AAKEAELANKQKEAKAKQDKTKDDQVADQQTVTVTTSQEKVTDKADTAQAQADLTA

Query Match 42.9%; Score 29.6; DB 1; Length 57660;
Best Local Similarity 64.7%; Pred. No. 20;
Matches 44; Conservative 0; Mismatches 24; Indels 0; Gaps 0;
Qy 2 ATGCTAATCTGTAGGAGTTTCAATTTTTCAGCAGGGTATCAGAACGGCTTTTACTCGTAATA 61

```

Db      56819 ATCGGCTTCTACAGATATATATTCATCAGACAGACAGTGAACGCATTACTTTAATA 56760
      ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Qy      62 TCATPACA 69
      | | | | |
Db      56759 TACCACAA 56752

RESULT 7
LOCUS   AC123142
DEFINITION Rattus norvegicus clone CH230-144K23, *** SEQUENCING IN PROGRESS
***, 5 unordered pieces.
ACCESSION AC123142.5 GI:30579048
VERSION   HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.
KEYWORDS  Rattus norvegicus (Norway rat)
SOURCE    Rattus norvegicus
ORGANISM  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 224280)
Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Albrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Chen,Z., Chu,J.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., D'Souza,L.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregioris,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hollnabe,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,S., Kelly,S., Kelly,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensheva,L., Louissegh,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mavhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Minet,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwackeleme,O., Okwunnu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfankuch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puzoz,N., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajz,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
TITLE
JOURNAL

```

```

REFERENCE
AUTHORS 2 (bases 1 to 224280)
TITLE   Worley,K.C.
JOURNAL Direct Submission
        Submitted (29-MAY-2002) Human Genome Sequencing Center, Department
        of Molecular and Human Genetics, Baylor College of Medicine, One
        Baylor Plaza, Houston, TX 77030, USA
REFERENCE
AUTHORS 3 (bases 1 to 224280)
TITLE   Rat Genome Sequencing Consortium.
JOURNAL Direct Submission
        Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
        of Molecular and Human Genetics, Baylor College of Medicine, One
        Baylor Plaza, Houston, TX 77030, USA
        On May 13, 2003 this sequence version replaced gi:24942321.
        The sequence in this assembly is a combination of BAC based reads
        and whole genome shotgun sequencing reads assembled using Atlas
        (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
        in the feature table below represents a scaffold in the Atlas
        assembly (a 'contig-scaffold'). Within each contig-scaffold,
        individual sequence contigs are ordered and oriented, and separated
        by sized gaps filled with Ns to the estimated size. The sequence
        may extend beyond the ends of the clone and there may be sequence
        contigs within a contig-scaffold that consist entirely of whole
        genome shotgun sequence reads. Both end sequences and whole genome
        shotgun sequence only contigs will be indicated in the feature
        table.
        ----- Genome Center
        Center: Baylor College of Medicine
        Center code: BCM
        Web site: http://www.hgsc.bcm.tmc.edu/
        Contact: hgsc-help@bcm.tmc.edu
        ----- Project Information
        Center project name: GMFV
        Center clone name: CH230-144K23
        ----- Summary Statistics
        Assembly program: Atlas 3.0;
        Consensus quality: 195444 bases at least Q40
        Consensus quality: 200343 bases at least Q30
        Consensus quality: 203534 bases at least Q20
        Estimated insert size: 208489; sum-of-contigs estimation
        Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
        -----
        * NOTE: Estimated insert size may differ from sequence length
        * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
        * NOTE: This is a 'working draft' sequence. It currently
        * consists of 5 contigs. The true order of the pieces
        * is not known and their order in this sequence record is
        * arbitrary. Gaps between the contigs are represented as
        * runs of N, but the exact sizes of the gaps are unknown.
        * This record will be updated with the finished sequence
        * as soon as it is available and the accession number will
        * be preserved
        * 1 12407: contig of 12407 bp in length
        * 12408 12507: gap of unknown length
        * 12508 109164: contig of 96657 bp in length
        * 109165 109264: gap of unknown length
        * 109265 168855: contig of 59591 bp in length
        * 168856 168955: gap of unknown length
        * 168956 22560: contig of 53605 bp in length
        * 22561 22560: gap of unknown length
        * 22561 224280: contig of 1620 bp in length.
        * 224280 124280: Location/Qualifiers
        * 1 124280
        * /organism="Rattus norvegicus"
        * /mol_type="genomic DNA"
        * /db_xref="taxon:10116"
        * /clones="CH230-144K23"
        * 94247. .96856
        * /note="wgs contig"
        * 121267. .123348
        * /note="wgs contig"
        *
        FEATURES
        source
        misc_feature
        misc_feature
        ORIGIN
        Query Match 40.6%; Score 28; DB 2; Length 224280;

```

```

Best Local Similarity 66.7%; Pred. No. 60;
Matches 40; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

Qy 3 TGCTAATCTGTAGGAGTTTCAATTTTCAGCAGGGTATCAGAACGGCTTTACTGGTAATAT 62
Db 148464 TACTAATCTGTGCCAGTAAATTTTCAGTACTACAGAACCCATTCACAGGAAGTAT 148523

RESULT 8
LOCUS BV162185/c 651 bp DNA linear STS 15-MAY-2004
DEFINITION RPAMSEQ0039680 Roche Palo Alto Mus musculus STS genomic, sequence
tagged site.
ACCESSION BV162185
VERSION BV162185.1 GI:47266772
KEYWORDS STS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 651)
AUTHORS Usuka,J., Liao,G., Cheng,J., Nguyen,A., Bach,C., Puech,A.,
McPherson,J.D., Foerzler,D. and Peltz,G.
Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 651)
Usuka,J., Liao,G., Cheng,J., Nguyen,A., Bach,C., Puech,A.,
McPherson,J.D., Foerzler,D. and Peltz,G.
Mus musculus SNPs
Unpublished (2003)

TITLE Contact: Jonathan Usuka
JOURNAL Roche Palo Alto Genetics and Genomics Department
COMMENT Roche Palo Alto
3431 Hillview Ave, Mailstop S3-1, Palo Alto, CA 94024, USA
Tel: 6508555807
Email: Jonathan.Usuka@roche.com
Primer A: No primer submitted with this STS
Primer B: No primer submitted with this STS.

FEATURES
source
1..651
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/map="7-6909-6270-CAAA01076479.1.1.9059"
/clone_lib="Roche Palo Alto"
/note="SNPs developed from assay sequences derived from 15
different strains of mice (as of October 1, 2003). Those
strains include A/J, A/HeJ, 129/Sv, AKR/J, B10.D2-H2/OsNj,
BALB/cByJ, BALB/cJ, C3H/HeJ, C57Bl/6J, CAST/Ei, DBA/2J,
MRL/MpJ, NZB/BinJ, NZW/LaC, SPRET/Ei."
<1..>651

STS
ORIGIN
Query Match 40.3%; Score 27.8; DB 11; Length 651;
Best Local Similarity 65.1%; Pred. No. 1.3e+02;
Matches 41; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

Qy 4 GCTAATCTGTAGGAGTTTCAATTTTCAGCAGGGTATCAGAACGGCTTTACTGGTAATATC 63
Db 285 GGTCACTAGGAAGAGTCTCAAACTGACCGCTGCCAAGATGGCTTAGCAGGTAAGAAAT 226

Qy 64 ACT 66
Db 225 GCT 223

RESULT 9
LOCUS AL139012 43551 bp DNA linear PRI 06-OCT-2001
DEFINITION Human DNA sequence from clone RP5-1091G18 on chromosome 1, complete
sequence.
ACCESSION AL139012
VERSION AL139012.14 GI:15990613
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 43551)
Garner,P.
Direct Submission
Submitted (06-OCT-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
requests: clonerequest@sanger.ac.uk
On Oct 9, 2001 this sequence version replaced gi:15958875.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em: EMBL; Sw:
SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/c_elegans/wormpep
This sequence
was generated from part of bacterial clone contigs of human
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chrl
RP5-1091G18 is from the library RPCI-5 constructed by the group of
Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pCYPAC2
IMPORTANT: This sequence is not the entire insert of clone
RP5-1091G18. It may be shorter because we sequence overlapping
sections only once, except for a short overlap.
The true left end of clone RP4-773A18 is at 41552 in this sequence.
The true right end of clone RP5-836N10 is at 2000 in this sequence.

FEATURES
source
1..43551
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP5-1091G18"
/clone_lib="RPCI-5"

ORIGIN
Query Match 40.0%; Score 27.6; DB 9; Length 43551;
Best Local Similarity 63.6%; Pred. No. 96;
Matches 42; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

Qy 3 TGCTAATCTGTAGGAGTTTCAATTTTCAGCAGGGTATCAGAACGGCTTTACTGGTAATAT 62
Db 23397 TGATCTATTCTAGGAACACACAGATTTAGCAGTGAATAAGACGTGCTTTTCATGCAATTT 23456

Qy 63 CACTAC 68
Db 23457 TTATAC 23462

RESULT 10
LOCUS AL135936/c 116341 bp DNA linear PRI 04-APR-2001
DEFINITION Human DNA sequence from clone RP11-3K6 on chromosome 20. Contains
STSS and GSSs, complete sequence.
ACCESSION AL135936
VERSION AL135936.11 GI:7378511
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 116341)
Matthews,L.
Direct Submission
Submitted (20-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
requests: clonerequest@sanger.ac.uk
On Mar 31, 2000 this sequence version replaced gi:7327706.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Emi, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 20, constructed by the Sanger Centre Chromosome 20
Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr20>
IMPORTANT: This sequence is not the entire insert of clone RP11-3K6
It may be shorter because we sequence overlapping sections only
once, except for a 100 base overlap.
The true left end of clone RP11-97N19 is at 116242 in this
sequence. The true right end of clone RP5-1068E13 is at 100 in this
sequence. This sequence was finished as follows unless otherwise
noted: all regions were either double-stranded or sequenced with an
alternate chemistry or covered by high quality data (i.e., phred
quality >= 30; an attempt was made to resolve all sequencing
problems, such as compressions and repeats; all regions were
covered by at least one plasmid subclone or more than one M13
subclone; and the assembly was confirmed by restriction digest.
RP11-3K6 is from the library RP11-11.1 constructed by the group of
Piter de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>
VECTOR: pBAC3.6.
Location/Qualifiers
1..116341
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="20"
/clone="RP11-3K6"
/clone_lib="RP11-11.1"
repeat_region
1738..1815
/note="MIR repeat: matches 71..144 of consensus"
repeat_region
2181..2687
/note="L1PA5 repeat: matches 5636..6143 of consensus"
repeat_region
4082..5621
/note="L1PB1 repeat: matches 4590..6155 of consensus"
repeat_region
5622..5907
/note="AluSg repeat: matches 1..286 of consensus"
repeat_region
5908..8448
/note="L1PB1 repeat: matches 2110..4590 of consensus"
repeat_region
8461..8784
/note="L1M2 repeat: matches 1676..2383 of consensus"
repeat_region
8794..15248
/note="L1PA7 repeat: matches 8..6141 of consensus"
repeat_region
15285..16327
/note="L1PBa repeat: matches -265..1692 of consensus"
repeat_region
16312..16861
/note="L1PBa repeat: matches -940..-401 of consensus"
repeat_region
16864..17064
/note="L1PBa repeat: matches -1543..-1343 of consensus"
repeat_region
17469..17843
/note="L1M4 repeat: matches 4017..4391 of consensus"
repeat_region
18629..18809
/note="M1A1 repeat: matches 324..541 of consensus"
repeat_region
18884..18927
/note="22 copies 2 mer ag 93% conserved"

misc_feature
19173..19685
/note="match: GSS: Em:AQ084711"
complement(20875..21293)
/note="match: GSS: Em:AQ039265"
complement(20920..21289)
/note="match: GSS: Em:AQ079510"
22013..22188
/note="MIR repeat: matches 17..203 of consensus"
22329..22683
/note="M1A1 repeat: matches 1..361 of consensus"
22911..27327
/note="L1HS repeat: matches 1731..6146 of consensus"
complement(27970..28271)
/note="match: GSS: Em:AZ006036"
complement(29311..29663)
/note="match: STS: Em:G18101"
29316..29443
/note="FLAM A repeat: matches 2..131 of consensus"
29680..29905
/note="MIR repeat: matches 14..249 of consensus"
30448..30570
/note="MER5A repeat: matches 1..127 of consensus"
30576..31434
/note="MER63C repeat: matches 1..922 of consensus"
31435..31494
/note="MER5A repeat: matches 129..189 of consensus"
31847..32585
/note="L2 repeat: matches 1931..2745 of consensus"
35126..35357
/note="116 copies 2 mer aa 63% conserved"
36129..36303
/note="AluSg/x repeat: matches 132..306 of consensus"
36498..36531
/note="17 copies 2 mer tg 100% conserved"
36532..36675
/note="AluSg/x repeat: matches 141..284 of consensus"
complement(36815..37554)
/note="match: GSS: Em:AQ345283"
36895..37663
/note="L1M3 repeat: matches 6956..7739 of consensus"
37664..37975
/note="AluJo repeat: matches 1..307 of consensus"
37976..38212
/note="L1M3 repeat: matches 6723..6956 of consensus"
38208..38445
/note="L1MD3 repeat: matches 6076..6329 of consensus"
38453..38888
/note="M1A1 repeat: matches 8..484 of consensus"
38889..39073
/note="M1A1-internal repeat: matches 2..181 of consensus"
39071..39285
/note="L1PA12 repeat: matches -62..153 of consensus"
39215..42394
/note="L1P repeat: matches 865..3997 of consensus"
42395..42506
/note="AluJo/FRAM repeat: matches 171..283 of consensus"
42542..42759
/note="M1A1 repeat: matches 106..305 of consensus"
42913..43150
/note="M1A1 repeat: matches 317..568 of consensus"
43152..43210
/note="MIR repeat: matches 22..80 of consensus"
43235..43266
/note="MER5A repeat: matches 43..74 of consensus"
44290..44445
/note="L2 repeat: matches 2338..2498 of consensus"
45489..45617
/note="M1A1 repeat: matches 79..211 of consensus"
45848..46336
/note="match: GSS: Em:AQ713111"
45852..46537
/note="match: GSS: Em:AQ392150"
47024..47340

repeat_region	/note="AluSx repeat: matches 1. .312 of consensus" 47486. .47619
repeat_region	/note="L1R16C repeat: matches 256. .387 of consensus" 48413. .48645
repeat_region	/note="L1MEC repeat: matches 2646. .2872 of consensus" 48466. .48867
repeat_region	/note="AluY repeat: matches 77. .301 of consensus" 48868. .50537
repeat_region	/note="L1MEC repeat: matches 1318. .2646 of consensus" 50570. .50850
repeat_region	/note="L1MEC repeat: matches 997. .1287 of consensus" 52742. .52932
repeat_region	/note="L1ME1 repeat: matches 5958. .6156 of consensus" 53300. .53550
repeat_region	/note="MIR repeat: matches 3. .260 of consensus" 53622. .53659
repeat_region	/note="19 copies 2 mer tt 81% conserved" 53664. .53866
repeat_region	/note="L1P8 repeat: matches 5955. .6157 of consensus" 54101. .54138
repeat_region	/note="19 copies 2 mer ac 84% conserved" 54763. .54894
repeat_region	/note="FLAM_C repeat: matches 2. .130 of consensus" 55956. .55999
repeat_region	/note="22 copies 2 mer gt 97% conserved" 56071. .56376
repeat_region	/note="AluSc repeat: matches 1. .306 of consensus" 58126. .58414
repeat_region	/note="L1R16C repeat: matches 81. .383 of consensus" 58985. .59100
repeat_region	/note="MIR repeat: matches 124. .242 of consensus" 60491. .60541
repeat_region	/note="M1R1G repeat: matches 494. .544 of consensus" 60542. .60701
repeat_region	/note="L1WA9 repeat: matches 6119. .6269 of consensus" 60711. .60776
repeat_region	/note="M1R1G repeat: matches 456. .529 of consensus" 60884. .61005
repeat_region	/note="MERSB repeat: matches 64. .178 of consensus" 61437. .61644
repeat_region	/note="L2 repeat: matches 2517. .2748 of consensus" 62123. .62189
repeat_region	/note="MIR repeat: matches 80. .145 of consensus" 62353. .62472
repeat_region	/note="FLAM_C repeat: matches 27. .143 of consensus"
Query Match	40.0%; Score 27.6; DB 9; Length 116341;
Best Local Similarity	72.0%; Pred.No. 87;
Matches 36; Conservative 0; Mismatches 14; Indels 0; Gaps 0;	
QY	1 GATGCTAATCTGTAGGAGTTCAATTTTCAGCAGGTTATCAGACGGCTT 50
Db	1168 GATGCTATTTCTATGGAGTTGTTTGTGTTGATTATCAGCAGGCTT 1119
RESULT 11	
AL357633/c	180298 bp DNA linear HTG 10-JUL-2001
LOCUS	Homo sapiens chromosome 1 clone RP11-542H20, 15 unordered pieces.
DEFINITION	AL357633
ACCESSION	AL357633
VERSION	AL357633.6 GI:11414582
KEYWORDS	HTG; HTGS PHASE1; HTGS_CANCELLED.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 Plumb, B.
TITLE	Direct Submission
JOURNAL	Submitted (10-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT	On Nov 28, 2000 this sequence version replaced qi:10241378.

```
misc_feature      fragment_chain:1"
25527..32211
/note="assembly_fragment:02409
fragment_chain:1"
misc_feature      32312..35664
/note="assembly_fragment:01045
fragment_chain:2"
35765..40972
/note="assembly_fragment:01398
fragment_chain:2"
41073..50462
/note="assembly_fragment:02008"
50563..68833
/note="assembly_fragment:01833
fragment_chain:3"
68934..71230
/note="assembly_fragment:00565
fragment_chain:3"
71331..90108
/note="assembly_fragment:02415
fragment_chain:3"
90209..103629
/note="assembly_fragment:01424
fragment_chain:3"
103730..110347
/note="assembly_fragment:01299
fragment_chain:3"
110448..137132
/note="assembly_fragment:00312
fragment_chain:3"
137233..151107
/note="assembly_fragment:01518
fragment_chain:3"
151208..180298
/note="assembly_fragment:02207
fragment_chain:3
clone_end:T7
vector_side:right"

ORIGIN
Query Match      40.0%; Score 27.6; DB 2; Length 180298;
Best Local Similarity 63.6%; Pred. No. 83;
Matches 42; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

Qy 3 TGCTAATCTAGGAGTTTCAATTCAGCAGGGTATCAGACGGCTTACTGGTATAT 62
Db 12969 TGATACTATTCTAGGACACAGATTTAGCAGTGAATAAGACATGCTTTTCATGGCAATTT 12910

Qy 63 CACTAC 68
Db 12909 TTATAC 12904

RESULT 12
BX530036 BX530036 211643 bp DNA linear VRT 27-SEP-2003
LOCUS Zebrafish DNA sequence from clone DKEY-93M18 in linkage group 21,
DEFINITION complete sequence.
ACCESSION BX530036
VERSION BX530036.7 GI:35209654
KEYWORDS HTG
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 211643)
AUTHORS Heath,P.
TITLE Direct Submission
JOURNAL Submitted (27-SEP-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zf1sh-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT On Sep 24, 2003 this sequence version replaced gi:33620376.
```

```
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zf1sh-help@sanger.ac.uk
-----
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest, except on the rare
occasion of the clone being a YAC.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep/Clone-derived
Zebrafish pUC subclones occasionally display inconsistency over the
length of mononucleotide A/T runs and conserved TA repeats. Where
this is found the longest good quality representation will be
submitted.
Repeat names beginning 'Dr' were identified by the Recon repeat
discovery system (Zhihong Bao and Sean Eddy, submitted), and those
beginning 'drr' were identified by Rick Waterman (Stephen Johnson
lab, WashU). For further information see
http://www.sanger.ac.uk/Projects/D_rerio/fishmask.shtml DKEY-93M18
is from a Zebrafish BAC library
VECTOR: pindigobac-5.
FEATURES             source
                Location/Qualifiers
                1..211643
                /organism="Danio rerio"
                /mol_type="genomic DNA"
                /db_xref="taxon:7955"
                /clones="DKEY-93M18"
                /clone_lib="DanioKey"

ORIGIN
Query Match      40.0%; Score 27.6; DB 5; Length 211643;
Best Local Similarity 63.6%; Pred. No. 82;
Matches 42; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

Qy 2 ATGCTAATCTAGGAGTTTCAATTCAGCAGGGTATCAGACGGCTTACTGGTATA 61
Db 90735 ATTTTAATCTTGTAGAAGAATCATTTTATCAGTGTTCAGAAATTAATTGATATTA 90794

Qy 62 TCACTA 67
Db 90795 ACAGTA 90800

RESULT 13
AC095519/c AC095519/c 225538 bp DNA linear HTG 09-MAY-2003
LOCUS Rattus norvegicus clone CH230-707, *** SEQUENCING IN PROGRESS ***,
DEFINITION 2 unordered pieces.
ACCESSION AC095519
VERSION AC095519.7 GI:30467460
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 225538)
AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
```

Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandarainaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswalo, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasari, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M., Davis, C., Davy-Carroll, L., De Angelis, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escoto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Frazer, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuhewa, L., Louissegh, H., Lozado, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhiney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokemele, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodkey, J., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umanai, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R., Woodend, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, J., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstein, G. and Gibbs, R.A.

TITLE
JOURNAL
REFERENCE
AUTHORS

2 (bases 1 to 225538)
Worley, K.C.
Direct Submission
Submitted (17-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

TITLE
JOURNAL
REFERENCE
AUTHORS

3 (bases 1 to 225538)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (09-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT

On May 9, 2003 this sequence version replaced gi:24940680.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome

shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
----- Project Information
Center project name: GCPE
Center clone name: CH230-707
----- Summary Statistics

Assembly program: Atlas;
Consensus quality: 181494 bases at least Q40
Consensus quality: 186363 bases at least Q30
Consensus quality: 189680 bases at least Q20
Estimated insert size: 189747; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 3592: contig of 3592 bp in length
* 3593 3692: gap of unknown length
* 3693 225538: contig of 221846 bp in length.

FEATURES

source

1..225538
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clones="CH230-707"

misc_feature

1..1227
/note="wgs contig"

misc_feature

2104..3592
/note="wgs contig"

misc_feature

58076..60333
/note="wgs contig"

misc_feature

74525..75949
/note="wgs contig"

misc_feature

complement(162735..221847)
/note="clone boundary"

misc_feature

clone_end:T7
site:EcoRI

misc_feature

end sequence:BH323723"
173486..175055

misc_feature

/note="wgs end_extension"

misc_feature

clone_end:T7"
175106..176422

misc_feature

/note="wgs end_extension"

ORIGIN

Query Match 40.0%; Score 27.6; DB 2; Length 225538;
Best Local Similarity 63.6%; Pred. No. 81;
Matches 42; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 3 TGCTAACTACTAGGAGTTTCAATTTACGACGGGTATCAGAACGCGCTTACTGGTAATAT 62
|||||
Db 62807 TGCTAGGTGGAGAGACTTCTAACTTCAGTGTGAAGCTAATGCTTTACAGGTACACA 62748
|||||
QY 63 CACTAC 68
|||||
Db 62747 CAGTAC 62742

RESULT 14
AC106652

LOCUS	AC106652	238937 bp	DNA	linear	HTG 13-MAY-2003
DEFINITION	Rattus norvegicus clone CH230-66P3, WORKING DRAFT SEQUENCE, 2 unordered pieces.				
ACCESSION	AC106652				
VERSION	AC106652.4	GI:30578574			
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.				
SOURCE	Rattus norvegicus				
ORGANISM	Rattus norvegicus (Norway rat)				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Ratus.				
REFERENCE	1 (bases 1 to 238937)				
AUTHORS	Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhaý,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Kocha,S., Dunn,A., Durbin,K., Duval,B., Evans,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,M., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpach,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovac,C., Kowts,C., Kraft,C.L., Lebow,H., Leván,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lourenshewa,L., Louisleged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Mirja,S., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokemeleh,O., Okwuonu,G., Olarnpunsaogoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puzoz,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartebyev,A., Sisason,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorrelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willison,R., Wlarczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Zuo,D., von Niederhausen,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.				
TITLE	Direct Submission				
REFERENCE	Unpublished				
AUTHORS	2 (bases 1 to 238937)				
TITLE	Worley,K.C.				
JOURNAL	Submitted (12-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA				
	3 (bases 1 to 238937)				
REFERENCE	Rat Genome Sequencing Consortium.				
AUTHORS	Submitted (13-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA				
TITLE	Direct Submission				
JOURNAL	Submitted (13-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA				
	1 GATCGTAATCTGTAGGAGTTTCAATTTCACGACGGGTATCATGACGCCGTTACTGTTAACT 60 Query Match 40.0%; Score 27.6; DB 2; Length 238937; Best Local Similarity 63.6%; Pred. No. 81; Matches 42; Conservative 0; Mismatches 24; Indels 0; Gaps 0;				
	of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA On May 13, 2003 this sequence version replaced gi:23264150. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequences are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.				
	----- Genome Center Center: Baylor College of Medicine Center code: BCM Web site: http://www.hgsc.bcm.tmc.edu/ Contact: hgsc-help@bcm.tmc.edu ----- Project Information Center project name: GLUTL Center Clone name: CH230-66P3 ----- Summary Statistics Assembly program: Atlas 3.0; Consensus quality: 227435 bases at least Q40 Consensus quality: 229962 bases at least Q30 Consensus quality: 231424 bases at least Q20 Estimated insert size: 237787; sum-of-contigs estimation Quality coverage: 7x in Q20 bases; sum-of-contigs estimation ----- * NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html). * NOTE: This is a 'working draft' sequence. It currently consists of 2 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved. * 1 237740: contig of 237740 bp in length * 237741 237840: gap of unknown length * 237841 238937: contig of 1097 bp in length. FEATURES Location/Qualifiers 1..238937 /organism="Rattus norvegicus" /mol_type="genomic DNA" /db_xref="taxon:10116" /clone="CH230-66P3" misc_feature 1..2692 /note="wgs end extension clone_end:T7" misc_feature 7349..8179 /note="clone boundary clone_end:T7" site:EcoRI end_sequence:BH269831" misc_feature 134429..135821 /note="wgs contig" complement(236160..236748) /note="clone boundary clone_end:Sp6 site:EcoRI end_sequence:BH269833"				
ORIGIN					
	Query Match 40.0%; Score 27.6; DB 2; Length 238937; Best Local Similarity 63.6%; Pred. No. 81; Matches 42; Conservative 0; Mismatches 24; Indels 0; Gaps 0;				
QY	1 GATCGTAATCTGTAGGAGTTTCAATTTCACGACGGGTATCATGACGCCGTTACTGTTAACT 60 				

Db 90745 GAGGTAAAGAGTAGGAGTTGTGATTGCTCAGTGTCCCAATTAAGGCTCTTTTGGGAAG 90804

Qy 61 ATCACT 66

Db 90805 AAAACT 90810

RESULT 15

AC105534

LOCUS

DEFINITION Rattus norvegicus clone CH230-22109, *** SEQUENCING IN PROGRESS

ACCESSION AC105534

VERSION AC105534.4 GI:24942038

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.

SOURCE Rattus norvegicus (Norway rat)

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

1 (bases 1 to 227763)

Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alabrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalto,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.B., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denison,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowitz,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseghe,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindaratne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhinney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokeme,O., Okwuonu,G., Olarpunsgoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfankoch,C., Plopper,F., Poindecker,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shateman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

Direct Submission

Unpublished

2 (bases 1 to 227763)

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

***** Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

***** Project Information

Center project name: GNDK

Center clone name: CH230-22109

***** Summary Statistics

Assembly program: Phrap; version 0.990129

Consensus quality: 190899 bases at least Q40

Consensus quality: 194726 bases at least Q30

Consensus quality: 197408 bases at least Q20

Estimated insert size: 194609; sum-of-contrigs estimation

Quality coverage: 6x in Q20 bases; sum-of-contrigs estimation

* NOTE: Estimated insert size may differ from sequence length

* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently

* consists of 3 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 223270: contig of 223270 bp in length

* 223271 223370: gap of unknown length

* 223371 226446: contig of 3076 bp in length

* 226447 226546: gap of unknown length

* 226547 227763: contig of 1217 bp in length.

***** Location/Qualifiers

1..227763

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/db_xref="taxon:10116"

/clone="CH230-22109"

1..1208

/notes="wgs contig"

6480..9433

/notes="wgs contig"

219496..221046

/notes="wgs contig"

223371..224889

/notes="wgs contig"

ORIGIN

Query Match 39.7%; Score 27.4; DB 2; Length 227763;

Best Local Similarity 69.8%; Pred. No. 94;

Matches 37; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy 2 ATGCTAATACTGTAGGAGTTTCAATTTCAGCAGGGTATCAGAACGGCTTTACT 54
Db 102670 ATGTTAATCTTCTTAGGAGTTAGAGTCCAGCAGTGAATCAGACAGGTAAGACT 102722

Search completed: November 20, 2005, 07:31:37
Job time : 1460 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 2005, 07:07:25 ; Search time 614 Seconds
(without alignments)
665.248 Million cell updates/sec

Title: US-10-780-250-4
Perfect score: 69
Sequence: 1 gatgctaatactgttagagt.....ttactggaataatacactaca 69

Scoring table: OLIGO NUC
Gapop 60.0 , Gapext 60.0

Searched: 4390206 seqs, 2959870667 residues

Word size : 0

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

- Database : N_Geneseq_16Dec04.*
- 1: Geneseqn1980s.*
 - 2: Geneseqn1990s.*
 - 3: Geneseqn2000s.*
 - 4: Geneseqn2001as.*
 - 5: Geneseqn2001bs.*
 - 6: Geneseqn2002as.*
 - 7: Geneseqn2002bs.*
 - 8: Geneseqn2003as.*
 - 9: Geneseqn2003bs.*
 - 10: Geneseqn2003cs.*
 - 11: Geneseqn2003ds.*
 - 12: Geneseqn2004as.*
 - 13: Geneseqn2004bs.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	18	26.1	7568	2 AAX27959	Sugarcane
C 2	18	26.1	7568	2 AAA73991	Aaa73991 Sugarcane
C 3	18	26.1	7568	2 AAZ39510	Aaz39510 Sugarcane
C 4	18	26.1	7568	8 ABX93826	Abx93826 Sugarcane
C 5	17	24.6	480	5 ADL38796	Adl38796 Human ova
C 6	17	24.6	481	5 ADI67577	Adi67577 Human ova
C 7	17	24.6	2016	13 ADS47069	Ads47069 Bacterial
C 8	17	24.6	3379	12 ADO39649	Ado39649 Yeast CHS
C 9	17	24.6	3625	13 ADT05481	Adt05481 Haemophil
C 10	17	24.6	6591	12 ADP04338	Adp04338 Squid pot
C 11	17	24.6	31129	6 AAD36229	Aad36229 Human tra
C 12	17	24.6	135356	13 ADT05646	Adt05646 Haemophil
C 13	16	23.2	326	12 ADL84829	Adl84829 DNA up-re
C 14	16	23.2	326	12 ADL84828	Adl84828 DNA up-re
C 15	16	23.2	326	12 ADL84830	Adl84830 DNA up-re
C 16	16	23.2	513	5 ABV58455	Abv58455 Human pro
C 17	16	23.2	597	13 ADL19228	Adl19228 Bacteriop
C 18	16	23.2	666	5 AAS71709	Aas71709 DNA encod
C 19	16	23.2	967	10 ADE81111	Ade81111 Bovine l-
C 20	16	23.2	1188	12 ADP98579	Adp98579 C. albica

C 21	16	23.2	1275	13 ADT46409	Adt46409 Bacterial
C 22	16	23.2	1830	8 AAD47827	Aad47827 Haemophil
C 23	16	23.2	1833	4 AAS53284	Aas53284 Haemophil
C 24	16	23.2	1833	8 ACA34109	Aca34109 Prokaryot
C 25	16	23.2	3090	6 ABZ14386	Abz14386 Arabidops
C 26	16	23.2	3942	2 AAX99550	Aax99550 Nucleic a
C 27	16	23.2	3942	8 ACAS2809	Acas2809 Prokaryot
C 28	16	23.2	8894	13 ADT05524	Adt05524 Haemophil
C 29	16	23.2	33393	13 ABD33223	Abd33223 Human can
C 30	16	23.2	36135	4 AAK84218	Aak84218 Human imm
C 31	16	23.2	96593	9 ADA02885	Ada02885 Mouse Blm
C 32	16	23.2	96593	10 ADB72623	Adb72623 Mouse Blm
C 33	16	23.2	96593	12 ADM74480	Adm74480 Murine ca
C 34	16	23.2	96594	10 ADC85364	Adc85364 Human Pap
C 35	16	23.2	110000	2 AAT42063	04 Continuation (5 of
C 36	16	23.2	117576	13 ADT05736	Adt05736 Haemophil
C 37	16	23.2	178825	11 ACN45144	Acn45144 Mouse gen
C 38	15	21.7	113	12 ADO13631	Ado13631 SNP targe
C 39	15	21.7	190	3 AAA41504	Aaa41504 Human sec
C 40	15	21.7	223	9 ADA09632	Ada09632 Restricti
C 41	15	21.7	248	3 AAA57037	Aaa57037 Human col
C 42	15	21.7	248	6 ABT12459	Abt12459 Orestes s
C 43	15	21.7	248	10 ACD91753	Acd91753 Human col
C 44	15	21.7	255	6 ADH31826	Adh31826 Yeast sma
C 45	15	21.7	413	8 ABX44290	Abx44290 Bovine ES

ALIGNMENTS

RESULT 1
AAX27959/c
ID AAX27959 standard; DNA; 7568 BP.

AC AAX27959;
XX
AC AAX27959;
DT 08-JUN-1999 (first entry)
XX
DE Sugarcane bacilliform virus genomic sequence.
XX
KW Sugarcane bacilliform virus; promoter; SCBV promoter; transgenic plant;
KW water deficit resistance; pest resistance; herbicide resistance;
KW nutritional content; protein manufacture; ss.

OS Sugarcane bacilliform virus.
XX
XX WO9909190-A1.
XX
PD 25-FEB-1999.
XX
PF 13-AUG-1997; 97WO-IB001338.
XX
PR 13-AUG-1997; 97WO-IB001338.

XX (MINU) UNIV MINNESOTA.
XX
XX Olszewski N, Tzafrir I, Somers DA, Lockhart B, Torbert K;
XX
XX WPI; 1999-181047/15.
XX
XX P-PSDB; AAY01078.
XX
XX Newly isolated DNA molecule comprising a sugarcane bacilliform virus
XX promoter - useful for generating fertile transgenic plants with
XX beneficial agronomic characteristics.
XX
XX Disclosure; Fig 2; 67pp; English.

XX This sequence represents a sugarcane bacilliform virus promoter. The
XX claims present in the specification refer to Seq ID Numbers that are
XX specific fragments of this sequence, however, there is no indication in
XX the specification as to which nucleotides of the promoter sequence
XX correspond to the specifically claimed regions. The invention relates to
XX a preselected DNA segment, consisting of a sugarcane bacilliform virus

CC (ScBV) promoter (including a biologically active subunit). Fertile
 CC transgenic plants produced by transformation with the DNA sequence are
 CC useful for conventional plant breeding programs. Other transgenic plants
 CC are useful for generating beneficial agronomic characteristics, including
 CC resistance to water deficit, pest resistance, herbicide resistance,
 CC increased yield, improved nutritional content, and improved processing
 CC characteristics. They are also useful for commercial protein manufacture.
 CC The ScBV promoter is a strong, constitutive promoter in may cells,
 CC enabling the expression of genes in fertile transgenic monocot or dicot
 CC plants, which allows genetic engineering of improved plant
 CC characteristics

SQ Sequence 7568 BP; 2602 A; 1472 C; 1805 G; 1689 T; 0 U; 0 Other;

Query Match 26.1%; Score 18; DB 2; Length 7568;

Best Local Similarity 100.0%; Pred. No. 7.9;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 21 TTCAATTTTCAGCAGGGTA 38

Db 2684 TTCAATTTTCAGCAGGGTA 2667

RESULT 2

AAA73991/c
 ID AAA73991 standard; DNA; 7568 BP.

XX AC AAA73991;

XX DT 01-DEC-2000 (first entry)

XX DE Sugarcane bacilliform virus nucleotide sequence.

XX KW Sugarcane bacilliform virus; ScBV; ScBV promoter; transgenic plant; ds.

XX OS Sugarcane bacilliform virus.

FH Key Location/Qualifiers
 FT CDS 1536..7151
 FT /*tag= a

XX PN US6093569-A.

XX PD 25-JUL-2000.

XX PF 08-JUL-1999; 99US-00349546.

XX PR 09-AUG-1996; 96US-00694869.

XX PR 13-AUG-1997; 97WO-IB001338.

XX PA (MINU) UNIV MINNESOTA.

XX PI Lockhart B, Torbert K, Olszewski N, Tzafrir I, Somers DA;

XX DR WPI; 1999-181047/15.

XX DR P-PSDB; AAB15671.

XX PT Newly isolated DNA molecule comprising a sugarcane bacilliform virus
 PT promoter - useful for generating fertile transgenic plants with
 PT beneficial agronomic characteristics.

XX PS Disclosure; Fig 2; 34pp; English.

XX CC The present sequence is the nucleotide sequence of the sugarcane
 CC bacilliform virus (ScBV). An ScBV gene promoter may be used to express
 CC proteins and RNA transcripts in transgenic plants. The promoter is
 CC operably linked to the desired nucleotide sequence and the recombinant
 CC DNA is introduced into plant cells. The method is useful for producing
 CC plants having unique signature sequences or other marker sequences which
 CC can be used to identify proprietary lines or varieties. Transgenic plants
 CC can be created for use in traditional agriculture and for use in the
 CC manufacture of proteins or other compounds, which are extracted or
 CC purified from plant parts or seeds. They may also be used in commercial

CC breeding programmes, including in the creation of new mutant plants
 CC through insertional mutagenesis

SQ Sequence 7568 BP; 2602 A; 1468 C; 1809 G; 1689 T; 0 U; 0 Other;

Query Match 26.1%; Score 18; DB 2; Length 7568;

Best Local Similarity 100.0%; Pred. No. 7.9;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 21 TTCAATTTTCAGCAGGGTA 38

Db 2684 TTCAATTTTCAGCAGGGTA 2667

RESULT 3

AAZ39510/c

ID AAZ39510 standard; DNA; 7568 BP.

XX AC AAZ39510;

XX DT 11-FEB-2000 (first entry)

XX DE Sugarcane bacilliform virus gene sequence.

XX KW Sugarcane bacilliform virus; ScBV; promoter; transgenic plant;
 KW monocot plant; dicot plant; ss.

XX OS Sugarcane bacilliform virus.

XX PN US5994123-A.

XX PD 30-NOV-1999.

XX PF 09-AUG-1996; 96US-00694869.

XX PR 13-AUG-1997; 97WO-IB001338.

XX PA (MINU) UNIV MINNESOTA.

XX PI Lockhart B, Torbert K, Olszewski N, Tzafrir I, Somers DA;

XX DR WPI; 1999-181047/15.

XX DR P-PSDB; AAY57165.

XX PT Newly isolated DNA molecule comprising a sugarcane bacilliform virus
 PT promoter - useful for generating fertile transgenic plants with
 PT beneficial agronomic characteristics.

XX PS Disclosure; Fig 2A-B; 34pp; English.

XX CC The invention provides sugarcane bacilliform virus (ScBV) promoters
 CC selected from sequences shown in AAZ39511-513. The promoter is used to
 CC express proteins and RNA transcripts in transgenic plants in order to
 CC improve their characteristics. The promoter confers constitutively high
 CC levels of expression of operably linked preselected DNA sequences in both
 CC monocot and dicot plants, plant tissue, plant parts or plant cells. The
 CC present sequence represents the sugarcane bacilliform virus gene sequence

SQ Sequence 7568 BP; 2602 A; 1468 C; 1809 G; 1689 T; 0 U; 0 Other;

Query Match 26.1%; Score 18; DB 2; Length 7568;

Best Local Similarity 100.0%; Pred. No. 7.9;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 21 TTCAATTTTCAGCAGGGTA 38

Db 2684 TTCAATTTTCAGCAGGGTA 2667

RESULT 4

ABX93826/c

ID ABX93826 standard; DNA; 7568 BP.

XX

AC ABX93826;
XX 03-JUN-2003 (first entry)
XX Sugarcane bacilliform virus (ScBV) genome.
XX
XX ScBv; ds; genome; promoter; transgenic plant; agronomic property;
XX stress tolerance; pest resistance; disease resistance; yield;
XX nutritional value; grain composition.
XX
XX Sugarcane bacilliform virus.
XX
XX Key Location/Qualifiers
XX CDS 1536..7151
XX /*tag= a
XX /product= "ScBV protein"
XX promoter 599..7420
XX /*tag= b
XX /note= "Claimed in claim 1"
XX
XX US6489462-B1.
XX
XX 03-DEC-2002.
XX
XX 11-FEB-2000; 2000US-00502831.
XX
XX 09-AUG-1996; 96US-00694869.
XX 13-AUG-1997; 97WO-IB001338.
XX 08-JUL-1999; 99US-00349546.
XX (MINU) UNIV MINNESOTA.
XX
XX Olaszewski N, Tzafrir I, Somers DA, Lockhart B, Torbert KA;
XX WPI; 2003-327258/31.
XX P-PSDB; ABU08544.
XX
XX New sugarcane bacilliform virus promoters, useful for obtaining
XX transgenic plants (e.g. oat) with improved properties, e.g. increased
XX stress tolerance, pest or disease resistance, improved yields or improved
XX nutritional value.
XX
XX Disclosure; Fig 1; 37pp; English.
XX
XX The invention relates to an isolated and purified DNA molecule, which
XX comprises a promoter (appearing as ABX93829, Promoter 1) or its fragment
XX with promoter activity, where the DNA molecule does not encode a
XX sugarcane bacilliform virus (ScBV) polypeptide. Also included is an
XX expression cassette comprising a first DNA segment, which comprise
XX Promoter1 or its fragment with promoter activity, functional in a plant
XX cell operably linked to a second DNA segment encoding a protein or RNA
XX transcript, where the second DNA segment does not encode a sugarcane
XX bacilliform virus polypeptide. The DNA or promoter is useful for
XX expressing proteins and/or RNA transcripts in transgenic plants, or for
XX conferring constitutively high levels of expression of operably linked
XX selected DNA segments in both monocotyledonous and dicotyledonous
XX plants, plant tissue, plant parts or plant cells. The DNA or promoter is
XX particularly useful in plant genetic engineering to obtain fertile
XX transgenic plants (e.g. oat, wheat, soybean, corn, tobacco, rice, barley,
XX potato, tomato, lettuce, oilseed rape, cotton, flax, sugar beet, sorghum,
XX sunflower, alfalfa, millet or rye) having improved agronomic properties,
XX characteristics or traits, e.g. increased stress tolerance, pest
XX resistance, disease resistance (e.g. bacteria, viruses or fungi),
XX improved yields, improved nutritional value, or improved grain
XX composition and quality. The present sequence is the ScBV genome from
XX which the promoters of the invention were isolated
XX
XX Sequence 7568 BP; 2602 A; 1468 C; 1809 G; 1689 T; 0 U; 0 Other;
XX
XX Query Match 26.1%; Score 18; DB 8; Length 7568;
XX Best Local Similarity 100.0%; Pred. No. 7.9;
XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 21 TTCAATTTTCAGCAGGTA 38
Db 2684 TTCAATTTTCAGCAGGTA 2667
RESULT 5
ADL38796/c
ID ADL38796 standard; DNA; 480 BP.
XX
XX AC ADL38796;
XX
XX DT 20-MAY-2004 (first entry)
XX
XX DE Human ovarian cancer DNA marker #12686.
XX
XX KW Human; ovarian cancer; ds; tumour; cytostatic; DNA marker.
XX
XX OS Homo sapiens.
XX
XX PN WO200170979-A2.
XX
XX PD 27-SEP-2001.
XX
XX PF 21-MAR-2001; 2001WO-US009126.
XX
XX PR 21-MAR-2000; 2000US-0191031P.
XX 25-MAY-2000; 2000US-0207124P.
XX 15-JUN-2000; 2000US-0211940P.
XX 07-JUL-2000; 2000US-0216820P.
XX 25-JUL-2000; 2000US-0220661P.
XX 21-DEC-2000; 2000US-0257672P.
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
XX Lee J, Lillie J;
XX WPI; 2001-611502/70.
XX
XX Novel isolated nucleic acid molecules (markers) overexpressed in ovarian
XX cancer cells as compared to their normal non-cancerous ovarian cells are
XX used to characterize stage, grade, histological type of ovarian cancer.
XX
XX Disclosure; SEQ ID NO 12686; 106pp; English.
XX
XX The invention relates to nucleic acid markers which are overexpressed in
XX ovarian cancer cells as compared to their expression in normal (i.e. non-
XX cancerous) ovarian cells. The invention also relates to polypeptides
XX encoded by the markers, antibodies that selectively bind to the
XX polypeptides, a method of inhibiting ovarian cancer in a patient at risk
XX of developing ovarian cancer involving inhibiting expression of a gene
XX corresponding to a marker of the invention and a method of treating a
XX patient afflicted with ovarian cancer comprising providing to cells of
XX the patient an antisense oligonucleotide complementary to a marker of the
XX invention. The markers are useful for assessing if a patient is afflicted
XX with ovarian cancer, which involves comparing the level of expression of
XX a marker in a patient sample and a normal level of expression of the
XX marker in a control non-ovarian cancer sample. A difference between the
XX expression levels indicates ovarian cancer. The level of expression of a
XX marker corresponds to a secreted protein or to a transcribed
XX polynucleotide or its portion. The level of expression of the marker is
XX assessed by detecting the presence in the sample, a protein or protein
XX fragment corresponding to the marker. The presence of protein or protein
XX fragment is detected using an antibody that specifically binds with the
XX protein or protein fragment. Alternatively, the level of expression of
XX the marker is assessed by detecting the presence of a transcribed
XX polynucleotide which anneals with the marker or anneals with a portion of
XX the polynucleotide comprising the marker, under stringent conditions. The
XX marker is also used for monitoring the progression of ovarian cancer in a
XX patient which involves detecting expression of the marker in a patient
XX sample at a first point in time, repeating the method at a subsequent
XX time and comparing the level of expression. The method is carried out
XX using an ovarian tissue sample. A composition comprising a marker,
XX polypeptide or antibody of the invention is used to treat ovarian cancer.

CC This sequence represents a human ovarian cancer DNA marker of the
CC invention.
XX
SQ Sequence 480 BP; 167 A; 58 C; 90 G; 164 T; 0 U; 1 Other;
Query Match 24.6%; Score 17; DB 5; Length 480;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 17 GAGTTTCAATTTCAGCA 33
|||||
Db 257 GAGTTTCAATTTCAGCA 241
RESULT 6
ADI67577/c
ID ADI67577 standard; DNA; 481 BP.
XX
AC ADI67577;
XX
DT 20-MAY-2004 (first entry)
XX
DE Human ovarian cancer DNA marker #319.
XX
KW Human; ovarian cancer; ds; tumour; cytostatic; DNA marker.
XX
OS Homo sapiens.
XX
PN WO200170979-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US009126.
XX
PR 21-MAR-2000; 2000US-0191031P.
XX
PR 25-MAY-2000; 2000US-0207124P.
XX
PR 15-JUN-2000; 2000US-0211940P.
XX
PR 07-JUL-2000; 2000US-0216820P.
XX
PR 25-JUL-2000; 2000US-0220661P.
XX
PR 21-DEC-2000; 2000US-0257672P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Lee J, Lillie J;
XX
DR WPI; 2001-611502/70.
XX
PT Novel isolated nucleic acid molecules (markers) overexpressed in ovarian
PT cancer cells as compared to their normal non-cancerous ovarian cells are
PT used to characterize stage, grade, histological type of ovarian cancer.
XX
FS Disclosure; SEQ ID NO 319; 106pp; English.
XX
CC The invention relates to nucleic acid markers which are overexpressed in
CC ovarian cancer cells as compared to their expression in normal (i.e. non-
CC cancerous) ovarian cells. The invention also relates to polypeptides
CC encoded by the markers, antibodies that selectively bind to the
CC polypeptides, a method of inhibiting ovarian cancer in a patient at risk
CC of developing ovarian cancer involving inhibiting expression of a gene
CC corresponding to a marker of the invention and a method of treating a
CC patient afflicted with ovarian cancer comprising providing to cells of
CC the patient an antisense oligonucleotide complementary to a marker of the
CC invention. The markers are useful for assessing if a patient is afflicted
CC with ovarian cancer, which involves comparing the level of expression of
CC a marker in a patient sample and a normal level of expression of the
CC marker in a control non-ovarian cancer sample. A difference between the
CC expression levels indicates ovarian cancer. The level of expression of a
CC marker corresponds to a secreted protein or to a transcribed
CC polynucleotide or its portion. The level of expression of the marker is
CC assessed by detecting the presence in the sample, a protein or protein
CC fragment corresponding to the marker. The presence of protein or protein
CC fragment is detected using an antibody that specifically binds with the
CC protein or protein fragment. Alternatively, the level of expression of

CC the marker is assessed by detecting the presence of a transcribed
CC polynucleotide which anneals with the marker or anneals with a portion of
CC the polynucleotide comprising the marker, under stringent conditions. The
CC marker is also used for monitoring the progression of ovarian cancer in a
CC patient which involves detecting expression of the marker in a patient
CC sample at a first point in time, repeating the method at a subsequent
CC time and comparing the level of expression. The method is carried out
CC using an ovarian tissue sample. A composition comprising a marker,
CC polypeptide or antibody of the invention is used to treat ovarian cancer.
CC This sequence represents a human ovarian cancer DNA marker of the
CC invention. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pt_sequences.
XX
SQ Sequence 481 BP; 168 A; 58 C; 90 G; 164 T; 0 U; 1 Other;
Query Match 24.6%; Score 17; DB 5; Length 481;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 17 GAGTTTCAATTTCAGCA 33
|||||
Db 257 GAGTTTCAATTTCAGCA 241
RESULT 7
ADS47069/c
ID ADS47069 standard; cDNA; 2016 BP.
XX
AC ADS47069;
XX
DT 02-DEC-2004 (first entry)
XX
DE Bacterial polynucleotide #1812.
XX
KW Recombinant DNA construct; transformed plant; improved plant property;
KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
KW pathogen tolerance; pest tolerance; plant disease resistance;
KW cell cycle pathway modification; plant growth regulator;
KW homologous recombination; seed oil yield; protein yield; carbohydrate;
KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
KW bacterial polynucleotide; gene; ss.
XX
OS Bacteria.
XX
PN US2003233675-A1.
XX
PD 18-DEC-2003.
XX
PF 20-FEB-2003; 2003US-00369493.
XX
PR 21-FEB-2002; 2002US-0360039P.
XX
PA (CAOY/) CAO Y.
PA (HINK/) HINKLE G J.
PA (SLAT/) SLATER S C.
PA (CHEN/) CHEN X.
XX (GOLD/) GOLDMAN B S.
PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
XX WPI; 2004-061375/06.
XX
PT New recombinant DNA construct comprising a promoter positioned to provide
PT for expression of a polynucleotide encoding a polypeptide from a
PT microbial source, useful for producing plants with improved properties.
XX
PS Claim 1; SEQ ID NO 25499; 122pp; English.
XX
CC The invention relates to a recombinant DNA construct comprising a
CC promoter functional in a plant cell, where the promoter is positioned to
CC provide for expression of a polynucleotide encoding a polypeptide from a
CC microbial source. The invention also relates to a transformed plant

CC comprising the recombinant DNA construct and a method of producing a
CC transformed plant having an improved property. The plant is a crop plant
CC such as maize or soybean. The method of producing a transformed plant
CC having an improved property comprises transforming a plant with the
CC recombinant DNA construct and growing the transformed plant, where the
CC polynucleotide or polypeptide is useful for improving plant properties.
CC The recombinant DNA construct is useful for producing plants with
CC improved plant properties, e.g. improved cold, heat or drought tolerance,
CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC increased resistance to plant disease, better growth rate by modification
CC of the cell cycle pathway with plant growth regulators, increased rate of
CC homologous recombination, modified seed oil or protein yield and/or
CC content, improved yield by modification of carbohydrate, nitrogen or
CC phosphorus use and/or uptake, by modification of photosynthesis or by
CC providing improved plant growth and development under at least one stress
CC condition. This sequence represents a bacterial polynucleotide used in
CC the scope of the invention. Note: The sequence data for this patent did
CC not form part of the printed specification but was obtained in electronic
CC format from USPTO at seqdata.uspto.gov/sequence.html.

XX SQ Sequence 2016 BP; 734 A; 381 C; 459 G; 442 T; 0 U; 0 Other;

Query Match 24.6%; Score 17; DB 13; Length 2016;

Best Local Similarity 100.0%; Pred. No. 28;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 GGAGTTTCAATTTCAGC 32

|||||

Db 1394 GGAGTTTCAATTTCAGC 1378

RESULT 8

AD039649/c

ID AD039649 standard; DNA; 3379 BP.

XX AC AD039649;

XX DT 29-JUL-2004 (first entry)

XX DE Yeast CHS5 (CAL3) gene.

XX KW Heterologous gene; expression cassette; gene expression; yeast; CHS5;

XX KW CAL3; gene; db.

XX OS Saccharomyces cerevisiae.

XX PN US2004092020-A1.

XX PD 13-MAY-2004.

XX PF 20-JUN-2003; 2003US-00600230.

XX PR 20-JUN-2002; 2002US-0390529P.

XX PA (WILK/) WILKINSON J Q.

XX PA (MCBR/) MCBRIDE K.

XX PA (BERT/) BERTAIN S.

XX PI Wilkinson JQ, McBride K, Bertain S;

XX DR WPI; 2004-374960/35.

XX DR GENBANK; 249198.

XX PT New recombinant expression cassette comprising a promoter that is
functional in plants, operably linked with a coding sequence and a non-
plant 3' termination sequence, useful for gene expression in plant cells.
XX PS Disclosure; SEQ ID NO 70; 74pp; English.
XX CC The present invention relates to heterologous genes comprising non-plant
3'-termination sequences and plant expression cassettes incorporating the
heterologous genes. The invention is useful for gene expression in plant

CC cells. The present sequence is yeast CHS5 (CAL3) gene. This sequence is
CC used in the invention.

XX SQ Sequence 3379 BP; 1156 A; 633 C; 694 G; 896 T; 0 U; 0 Other;

Query Match 24.6%; Score 17; DB 12; Length 3379;

Best Local Similarity 100.0%; Pred. No. 28;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 GGAGTTTCAATTTCAGC 32

|||||

Db 1879 GGAGTTTCAATTTCAGC 1863

RESULT 9

ADT05481/c

ID ADT05481 standard; DNA; 3625 BP.

XX AC ADT05481;

XX DT 02-DEC-2004 (first entry)

XX DE Haemophilus influenzae (NTHi) contig DNA sequence - SEQ ID 517.

XX KW middle ear bacterial infection; nasopharynx bacterial infection; db;

XX KW contig.

XX OS Haemophilus influenzae.

XX PN WO2004078949-A2.

XX PD 16-SEP-2004.

XX PF 05-MAR-2004; 2004WO-US007001.

XX PR 06-MAR-2003; 2003US-0453134P.

XX PA (CHIL-) CHILDRENS HOSPITAL INC.

XX PI Bakaletz LO, Munson RS, Dyer DW;

XX DR WPI; 2004-662422/64.

XX PT New polynucleotides of nontypeable strain of Haemophilus influenzae,
useful for treating or preventing NTHi bacterial infections of the middle
ear and/or nasopharynx.

XX PS Example 1; SEQ ID NO 517; 88pp; English.

XX CC The invention comprises nucleotide sequences (genes) from the genome of a
nontypeable strain of Haemophilus influenzae (NTHi). The NTHi DNA
sequences of the invention are useful for treating or preventing NTHi
bacterial infections of the middle ear and/or nasopharynx. The present
nucleic acid represents an NTHi contig sequence of the invention.

XX SQ Sequence 3625 BP; 1150 A; 626 C; 730 G; 1116 T; 0 U; 3 Other;

Query Match 24.6%; Score 17; DB 13; Length 3625;

Best Local Similarity 100.0%; Pred. No. 28;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 GAGTTTCAATTTCAGCA 33

|||||

Db 2796 GAGTTTCAATTTCAGCA 2780

RESULT 10

ADP04338

ID ADP04338 standard; DNA; 6591 BP.

XX AC ADP04338;

XX DT 29-JUL-2004 (first entry)

XX Squid potential dependent calcium channel-related gene #1.
DE screening; invertebrate animal potential dependent calcium channel;
XX potential-dependent calcium channel alpha 1 subunit; LoCav2;
KW potential-dependent calcium-channel beta subunit;
KW myogenic potential-dependent calcium channel alpha-1 subunit;
KW myogenic potential-dependent calcium channel delta subunit;
KW insect repellent; pesticide; squid; gene; ds.
XX Loligo bleekeri.
OS Location/Qualifiers
XX Key 1..6591
FH CDS
FT /product= "Squid potential dependent calcium channel-
FT related protein #1"
FT /transl_except= (pos:4594..4596, aa:Lys)
XX JP2004125786-A.
PN 22-APR-2004.
XX 29-AUG-2003; 2003JP-00307622.
XX 29-AUG-2002; 2002JP-00251885.
PR (DOKU-) DOKURITSU GYOSEI HOJIN SANGYO GIJUTSU SO.
PA WPI; 2004-381742/36.
DR P-PSDB; ADP04337.
XX Material useful for screening agonist of invertebrate animal potential
PT dependent calcium channel, comprises spear squid potential-dependent-
PT calcium-channel alpha 1 sub-unit (LoCav2) or its variant protein.
XX Disclosure; SEQ ID NO 2; 14pp; Japanese.
XX The invention comprises a method of screening an agonist of invertebrate
CC animal potential dependent calcium channel. The method involves spear
CC squid potential-dependent calcium channel alpha 1 subunit (LoCav2),
CC potential-dependent-calcium-channel beta subunit, and mammal myogenic
CC potential-dependent calcium channel alpha-1/delta subunit. The method of
CC the invention is useful for screening an agonist of invertebrate animal
CC potential-dependent calcium channel, and contributes in research of an
CC insect repellent such as a pesticide. The present DNA sequence encodes a
CC squid protein that was used in the exemplification of the invention.
XX SQ Sequence 6591 BP; 1948 A; 1303 C; 1433 G; 1906 T; 0 U; 1 Other;
Query Match 24.6%; Score 17; DB 12; Length 6591;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 18 AGTTTCAATTTTCAGCAG 34
|||||
DB 5129 AGTTTCAATTTTCAGCAG 5145
RESULT 11
AAD36229
ID AAD36229 standard; DNA; 31129 BP.
XX AAD36229;
AC
XX 09-AUG-2002 (first entry)
DT
XX Human transporter genomic DNA.
DE
XX Human; transporter protein; immune response; pharmacogenomic analysis;
KW drug screening; gene therapy; single nucleotide polymorphism; SNP;
KW immunostimulant; chromosome 18; gene; ds.
XX

OS Homo sapiens.
XX Key
FH variation
FT /tag= a
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace(11751, A)
FT /tag= b
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace(11944, T)
FT /tag= c
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT 12789..30395
FT /tag= d
FT /product= "Human transporter protein"
FT 12789..12938
FT /tag= e
FT 12939..13806
FT /tag= f
FT 13807..13987
FT /tag= g
FT 13988..15081
FT /tag= h
FT 15082..15271
FT /tag= i
FT 15272..16924
FT /tag= j
FT replace(16706, -)
FT /tag= k
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT 16925..17053
FT /tag= l
FT 17054..20897
FT /tag= m
FT replace(17269, T)
FT /tag= n
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace(17548, A)
FT /tag= o
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace(19904, G)
FT /tag= p
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT 20898..21090
FT /tag= q
FT 21091..23654
FT /tag= r
FT 23655..23802
FT /tag= s
FT 23803..25118
FT /tag= t
FT replace(24723, G)
FT /tag= u
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT 25119..25253
FT /tag= v
FT 25254..27043
FT /tag= w
FT 27044..27090
FT /tag= x
FT 27091..30212
FT /tag= y
FT replace(27527, G)
FT /tag= z
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace(28024, C)
FT /tag= aa
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace(28335, G)
FT /tag= ab
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace(28789, T)
FT /tag= ac

```
FT variation /standard name= "Single nucleotide polymorphism (SNP) "  
FT replace(28987, T)  
FT /*tag= ad  
FT /standard name= "Single nucleotide polymorphism (SNP) "  
FT replace(29500, C)  
FT /*tag= ae  
FT /standard name= "Single nucleotide polymorphism (SNP) "  
FT 30213. .30395  
FT /*tag= ad  
XX  
XX WO200220763-A2.  
XX  
XX PD 14-MAR-2002.  
XX  
XX PF 07-SEP-2001; 2001WO-US028222.  
XX  
XX PR 07-SEP-2000; 2000US-0230702P.  
XX PR 10-OCT-2000; 2000US-00684390.  
XX  
XX PA (PEKE ) PE CORP NY.  
XX  
XX PI Ketchum KA, Difrancesco V, Beasley EM;  
XX  
XX DR WPI; 2002-393844/42.  
XX P-P5DB; AAE22853.  
XX  
XX PT New human transporter peptides and nucleic acids, useful in identifying  
XX PT modulators of transporter peptides, in pharmacogenomic analysis, or as  
XX PT targets for diagnosing a disease or predisposition to a disease mediated  
XX PT by the peptide.  
XX  
XX PS Claim 4; Fig 3; 90pp; English.  
XX  
XX CC The present invention relates to novel human transporter proteins and  
XX CC nucleic acids encoding them. Sequences of the invention are useful as  
XX CC query sequences in database searches to identify other family members or  
XX CC related sequences. The proteins may be used to raise antibodies or to  
XX CC elicit immune responses, in drug screening assays, as reagents in assays  
XX CC designed to quantitatively determine levels of the protein in biological  
XX CC fluids, as markers for tissues in which the corresponding protein is  
XX CC preferentially expressed, in cell-based or cell-free systems, to identify  
XX CC modulators of protein activity, as targets for diagnosing a disease or  
XX CC predisposition to a disease mediated by the peptide, in pharmacogenomic  
XX CC analysis, and to treat disorders characterised by inappropriate or  
XX CC unwanted protein expression. Nucleic acid sequences of the invention are  
XX CC useful in detecting mutations in transporter genes and gene expression  
XX CC product, as antisense constructs for controlling transporter gene  
XX CC expression in cells, tissues and organisms and in gene therapy of  
XX CC patients with cells that are aberrant in transporter gene expression. The  
XX CC present sequence is human transporter genomic DNA which is located on  
XX CC chromosome 18  
XX  
XX SQ Sequence 31129 BP; 8602 A; 6993 C; 7156 G; 8128 T; 0 U; 250 Other;  
XX  
XX Query Match 24.6%; Score 17; DB 6; Length 31129;  
XX Best Local Similarity 100.0%; Pred. No. 27;  
XX Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
XX Qy 28 TCAGCAGGGTATCAGAA 44  
XX |||||||||||||||  
XX Db 21215 TCAGCAGGGTATCAGAA 21231  
XX  
XX RESULT 12  
XX ADT05646/C  
XX ID ADT05646 standard; DNA; 135356 BP.  
XX  
XX AC ADT05646;  
XX  
XX DT 02-DEC-2004 (first entry)  
XX  
XX DE Haemophilus influenzae (NTHi) DNA sequence - SEQ ID 682.  
XX  
XX PT Classifying an unknown multi-lineage affiliated gene comprises isolating  
XX PT expressed nucleic acid sequences from the discrete cell sub-populations.  
XX
```

```
KW middle ear bacterial infection; nasopharynx bacterial infection; ds.  
XX  
XX OS Haemophilus influenzae.  
XX  
XX PN WO2004078949-A2.  
XX  
XX PD 16-SEP-2004.  
XX  
XX PF 05-MAR-2004; 2004WO-US007001.  
XX  
XX PR 06-MAR-2003; 2003US-0453134P.  
XX  
XX PA (CHIL-) CHILDRENS HOSPITAL INC.  
XX  
XX PI Bakaletz LO, Munson RS, Dyer DW;  
XX  
XX DR WPI; 2004-662422/64.  
XX  
XX PT New polynucleotides of nontypeable strain of Haemophilus influenzae,  
XX PT useful for treating or preventing NTHi bacterial infections of the middle  
XX PT ear and/or nasopharynx.  
XX  
XX PS Claim 1; SEQ ID NO 682; 88pp; English.  
XX  
XX CC The invention comprises nucleotide sequences (genes) from the genome of a  
XX CC nontypeable strain of Haemophilus influenzae (NTHi). The NTHi DNA  
XX CC sequences of the invention are useful for treating or preventing NTHi  
XX CC bacterial infections of the middle ear and/or nasopharynx. The present  
XX CC nucleic acid represents an NTHi DNA sequence of the invention.  
XX  
XX SQ Sequence 135356 BP; 41426 A; 24873 C; 27795 G; 41257 T; 0 U; 5 Other;  
XX  
XX Query Match 24.6%; Score 17; DB 13; Length 135356;  
XX Best Local Similarity 100.0%; Pred. No. 27;  
XX Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
XX Qy 17 GAGTTTCAATTCAGCA 33  
XX |||||||||||||||  
XX Db 134527 GAGTTTCAATTCAGCA 134511  
XX  
XX RESULT 13  
XX ADL84829/C  
XX ID ADL84829 standard; DNA; 326 BP.  
XX  
XX AC ADL84829;  
XX  
XX DT 20-MAY-2004 (first entry)  
XX  
XX DE DNA up-regulated in murine common lymphoid myeloid cells SeqID 1222.  
XX  
XX KW gene potential; multi-lineage; cell commitment; haematopoietic stem cell;  
XX KW HSC; multipotent progenitor; MPP; common lymphoid progenitor; CLP;  
XX KW common myeloid progenitor; CMP; bone marrow stem cell; mouse; murine; ds.  
XX  
XX OS Mus sp.  
XX  
XX PN WO2003093445-A2.  
XX  
XX PD 13-NOV-2003.  
XX  
XX PF 05-MAY-2003; 2003WO-US014114.  
XX  
XX PR 03-MAY-2002; 2002US-0377383P.  
XX  
XX PA (STOW-) STOWERS INST MEDICAL RES.  
XX  
XX PI Li L;  
XX  
XX DR WPI; 2004-022656/02.  
XX  
XX PT Classifying an unknown multi-lineage affiliated gene comprises isolating  
XX PT expressed nucleic acid sequences from the discrete cell sub-populations.  
XX
```

XX Claim 9; SEQ ID NO 1222; 123pp; English.

PS This invention relates to a novel method for predicting gene potential by

XX associating nucleic acid sequences of unknown function with particular

CC sub-population profiles. Specifically, it refers to classifying an

CC unknown multi-lineage affiliated gene by collecting hybridisation data to

CC develop a gene expression map, in order to determine the discrete sub-

CC population where it is expressed. The present invention describes methods

CC for predicting the lineage commitment of genes associated with the self-

CC renewing haematopoietic (blood) stem cells (HSCs), as well as the non-

CC self renewing multipotent progenitors (MPPs), common lymphoid progenitors

CC (CLPs) and common myeloid progenitors (CMPs), which are collectively

CC referred to as bone marrow stem cells populations. As such, these methods

CC can be used to identify associated multi-lineage affiliated genes and

CC hence the underlying molecular mechanisms in physiological haematopoietic

CC development. This polynucleotide sequence is DNA associated with a murine

CC CMP sub population of cells of the invention.

XX SQ Sequence 326 BP; 97 A; 48 C; 61 G; 116 T; 0 U; 4 Other;

Query Match 23.2%; Score 16; DB 12; Length 326;

Best Local Similarity 100.0%; Pred. No. 99;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 ATGCTAATCTGTAGG 17

Db 139 ATGCTAATCTGTAGG 124

RESULT 14

ADL84828/c

ID ADL84828 standard; DNA; 326 BP.

AC ADL84828;

XX 20-MAY-2004 (first entry)

XX DNA up-regulated in murine common lymphoid myeloid cells SeqID 1221.

DE gene potential; multi-lineage; cell commitment; haematopoietic stem cell;

XX HSC; multipotent progenitor; MPP; common lymphoid progenitor; CLP;

KW common myeloid progenitor; CMP; bone marrow stem cell; mouse; murine; ds.

XX Mus sp.

OS WO2003093445-A2.

PN 13-NOV-2003.

PD 05-MAY-2003; 2003WO-US014114.

XX 03-MAY-2002; 2002US-0377383P.

PR (STOW-) STOWERS INST MEDICAL RES.

PA Li L;

XX WPI; 2004-022656/02.

DR Classifying an unknown multi-lineage affiliated gene comprises isolating

PT expressed nucleic acid sequences from the discrete cell sub-populations.

XX Claim 9; SEQ ID NO 1221; 123pp; English.

PS This invention relates to a novel method for predicting gene potential by

CC associating nucleic acid sequences of unknown function with particular

CC sub-population profiles. Specifically, it refers to classifying an

CC unknown multi-lineage affiliated gene by collecting hybridisation data to

CC develop a gene expression map, in order to determine the discrete sub-

CC population where it is expressed. The present invention describes methods

CC for predicting the lineage commitment of genes associated with the self-

CC renewing haematopoietic (blood) stem cells (HSCs), as well as the non-

CC self renewing multipotent progenitors (MPPs), common lymphoid progenitors

CC (CLPs) and common myeloid progenitors (CMPs), which are collectively

CC referred to as bone marrow stem cells populations. As such, these methods

CC can be used to identify associated multi-lineage affiliated genes and

CC hence the underlying molecular mechanisms in physiological haematopoietic

CC development. This polynucleotide sequence is DNA associated with a murine

CC CMP sub population of cells of the invention.

XX SQ Sequence 326 BP; 97 A; 48 C; 61 G; 116 T; 0 U; 4 Other;

Query Match 23.2%; Score 16; DB 12; Length 326;

Best Local Similarity 100.0%; Pred. No. 99;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 ATGCTAATCTGTAGG 17

Db 139 ATGCTAATCTGTAGG 124

RESULT 15

ADL84830/c

ID ADL84830 standard; DNA; 326 BP.

XX AC ADL84830;

XX 20-MAY-2004 (first entry)

XX DNA up-regulated in murine common lymphoid myeloid cells SeqID 1223.

DE gene potential; multi-lineage; cell commitment; haematopoietic stem cell;

XX HSC; multipotent progenitor; MPP; common lymphoid progenitor; CLP;

KW common myeloid progenitor; CMP; bone marrow stem cell; mouse; murine; ds.

XX Mus sp.

OS WO2003093445-A2.

PN 13-NOV-2003.

PD 05-MAY-2003; 2003WO-US014114.

XX 03-MAY-2002; 2002US-0377383P.

PR (STOW-) STOWERS INST MEDICAL RES.

PA Li L;

XX WPI; 2004-022656/02.

DR Classifying an unknown multi-lineage affiliated gene comprises isolating

PT expressed nucleic acid sequences from the discrete cell sub-populations.

XX Claim 9; SEQ ID NO 1223; 123pp; English.

PS This invention relates to a novel method for predicting gene potential by

CC associating nucleic acid sequences of unknown function with particular

CC sub-population profiles. Specifically, it refers to classifying an

CC unknown multi-lineage affiliated gene by collecting hybridisation data to

CC develop a gene expression map, in order to determine the discrete sub-

CC population where it is expressed. The present invention describes methods

CC for predicting the lineage commitment of genes associated with the self-

CC renewing haematopoietic (blood) stem cells (HSCs), as well as the non-

CC self renewing multipotent progenitors (MPPs), common lymphoid progenitors

CC (CLPs) and common myeloid progenitors (CMPs), which are collectively

CC referred to as bone marrow stem cells populations. As such, these methods

CC can be used to identify associated multi-lineage affiliated genes and

CC hence the underlying molecular mechanisms in physiological haematopoietic

CC development. This polynucleotide sequence is DNA associated with a murine

CC CMP sub population of cells of the invention.

XX SQ Sequence 326 BP; 97 A; 48 C; 61 G; 116 T; 0 U; 4 Other;

Query Match 23.2%; Score 16; DB 12; Length 326;

Best Local Similarity 100.0%; Pred. No. 99;
Matches 16; Conservative 0; Mismatches 0;
Qy 2 ATGCTAATACTGTAGG 17
Db 139 ATGCTAATACTGTAGG 124

0; Indels 0; Gaps 0;

Search completed: November 20, 2005, 08:19:55
Job time : 625 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 2005, 04:47:40 ; Search time 279 Seconds
(without alignments)
1464.022 Million cell updates/sec

Title: US-10-780-250-4

Perfect score: 69

Sequence: 1 gatgctaatacttagtagt.....ttactggtaatactcataca 69

Scoring table: IDENTITY_NUC

Gapop 10_0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_16Dec04:*

1: Geneseqn1980s:*

2: Geneseqn1990s:*

3: Geneseqn2000s:*

4: Geneseqn2001as:*

5: Geneseqn2001bs:*

6: Geneseqn2002as:*

7: Geneseqn2002bs:*

8: Geneseqn2003as:*

9: Geneseqn2003bs:*

10: Geneseqn2003cs:*

11: Geneseqn2003ds:*

12: Geneseqn2004as:*

13: Geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	31.2	45.2	807	ADH83346	Adh83346 Enterococ
C 2	28.2	40.9	3211	ACA47564	ACA47564 Prokaryot
C 3	26.8	38.8	127917	ADR52731	Adr52731 Drug ther
C 4	26.4	38.3	1533	AAC51660	Aac51660 Arabidops
C 5	26.4	38.3	1707	AAC46785	Aac46785 Arabidops
C 6	26.2	38.0	5978	ADQ63861	Adq63861 Novel hum
C 7	25.8	37.4	845	ADS60721	Ads60721 Bacterial
C 8	25.6	37.1	2379	ACA27631	ACA27631 Prokaryot
C 9	25.4	36.8	51952	AAV26084	Avv26084 Tomato pe
C 10	25.4	36.8	66681	ABD33333	Abd33333 Murine ca
C 11	25	36.2	322	ABN21249	Abn21249 Human ORF
C 12	25	36.2	352	ADQ20849	Adq20849 Human sof
C 13	25	36.2	500	AAA65428	Aaa65428 Porcine B
C 14	25	36.2	596	ADQ24808	Adq24808 Human sof
C 15	25	36.2	1308	ACC46307	Acc46307 Human dit
C 16	25	36.2	3814	ADJ56277	Adj56277 Human cdn
C 17	24.8	35.9	415	ACH15841	Ach15841 Human adu
C 18	24.8	35.9	3303	ABL03599	Ab103599 Drosophil
C 19	24.8	35.9	79084	ADQ97563	Adq97563 Human can
C 20	24.6	35.7	1030	ADS61254	Ads61254 Bacterial

21	24.4	35.4	667	10	ACF67221	Acf67221 Photorhab
22	24.4	35.4	780	10	ACF67010	Acf67010 Photorhab
23	24.4	35.4	1032	10	ACF69663	Acf69663 Photorhab
24	24.4	35.4	8372	4	ABL23536	Ab123536 Drosophil
25	24.4	35.4	34264	8	AA156706	Aa156706 Cynomolgu
C 26	24.4	35.4	110000	10	ACF67367_26	Continuation (27 o
27	24.4	35.4	110000	10	ACF65386_2	Continuation (3 of
28	24.4	35.4	110000	13	ABD32627_3	Continuation (4 of
C 29	24.2	35.1	722	5	AAS34119	Aas34119 Human cdn
C 30	24.2	35.1	738	6	ABN70831	Abn70831 Streptoco
C 31	24.2	35.1	783	6	ABN67812	Abn67812 Streptoco
C 32	24.2	35.1	783	6	ABN70413	Abn70413 Streptoco
C 33	24.2	35.1	1099	10	ADAL1386	Adal1386 Human int
C 34	24.2	35.1	1134	13	ADS46739	Ads46739 Bacterial
C 35	24.2	35.1	1262	10	ADC13562	Adc13562 Human NOV
C 36	24.2	35.1	1276	8	ABX72187	Abx72187 Human NOV
C 37	24.2	35.1	1463	8	ABX72186	Abx72186 Human NOV
C 38	24.2	35.1	1961	10	ADC30717	Adc30717 Human nov
C 39	24.2	35.1	8451	6	ABK39982	Abk39982 Human che
40	24.2	35.1	8451	6	ABL32659	Ab132659 Human imm
41	24.2	35.1	8451	6	AAS63319	Aas63319 Chemicall
C 42	24.2	35.1	15246	8	ABT13955	Abt13955 Simian Vi
C 43	24.2	35.1	22655	4	AAK70122	Aak70122 Human imm
44	24.2	35.1	110000	6	ABN71527_03	Continuation (4 of
45	24	34.8	433	6	ABA01732	Aba01732 Apple chl

ALIGNMENTS

RESULT 1

ADH83346/c
ID ADH83346 standard; DNA; 807 BP.

XX AC ADH83346;

XX DT 22-APR-2004 (first entry)

XX DE Enterococcus faecalis polynucleotide #1231.

XX KW Enterococcus faecalis infection; transcription regulatory element;
antibacterial; gene; db.

XX OS Enterococcus faecalis.

XX PN US6617156-B1.

XX PD 09-SEP-2003.

XX PF 13-AUG-1998; 98US-00134000.

XX PR 15-AUG-1997; 97US-0055778P.

XX PA (DOUC/) DOUCETTE-STAMM L A.
(BUSH/) BUSH D.

XX PI Doucette-Stamm LA, Bush D;

XX DR WPI; 2003-895394/82.

XX DR P-PSDB; ADH86751.

XX PT New nucleic acid comprising a sequence encoding an Enterococcus faecalis polypeptide, useful for preparing a composition for diagnosing or treating E. faecalis infection.

XX PS Disclosure; SEQ ID NO 1231; 193pp; English.

XX CC The invention relates to Enterococcus faecalis polynucleotides and polypeptides. The invention also relates to a recombinant expression vector comprising a polynucleotide operably linked to a transcription regulatory element, a cell comprising a recombinant vector, a method for producing an E. faecalis polypeptide, an isolated nucleic acid comprising a sequence not given in the specification, a recombinant vector

CC comprising the nucleic acid and a cell comprising the recombinant vector.
CC The polynucleotides can be used to detect the presence of E. faecalis in
CC a sample. The sequences are useful for preparing a composition for
CC diagnosing or treating Enterococcus faecalis infection. This sequence
CC represents an E. faecalis polynucleotide of the invention.
XX

SQ Sequence 807 BP; 295 A; 130 C; 145 G; 237 T; 0 U; 0 Other;

Query Match 45.2%; Score 31.2; DB 10; Length 807;

Best Local Similarity 66.2%; Pred. NO. 0.24;

Matches 45; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

Qy 2 ATGCTAATCTAGGAGTTTCAATTTTCAGCAGGATATCAGACGGCTTTACTGTAATA 61

Db 475 ATGCGGCTTACAGGATAATAATTTTCATCAGACAGTGAACGCAATTTACTGTAATA 416

Qy 62 TCACCTACA 69

Db 415 TACCTAAA 408

RESULT 2

ACA47564

ID ACA47564 standard; DNA; 3211 BP.

XX

AC ACA47564;

XX

DT 19-JUN-2003 (first entry)

XX

DE Prokaryotic essential gene #2921.

XX

KW Antisense; ds; prokaryotic essential gene; cell proliferation;

XX

KW drug design; gene.

XX

OS Staphylococcus haemolyticus.

XX

PN WO200277183-A2.

XX

PD 03-OCT-2002.

XX

PF 21-MAR-2002; 2002WO-US009107.

XX

PR 21-MAR-2001; 2001US-00815242.

XX

PR 06-SEP-2001; 2001US-00948993.

XX

PR 25-OCT-2001; 2001US-0342923P.

XX

PR 08-FEB-2002; 2002US-00072851.

XX

PR 06-MAR-2002; 2002US-0362699P.

XX

PA (ELIT-) ELITRA PHARM INC.

XX

PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

XX

PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX

DR WPI; 2003-029926/02.

XX

DR P-PSDB; ABU43694.

XX

XX

PT New antisense nucleic acids, useful for identifying proteins or screening

XX

PT for homologous nucleic acids required for cellular proliferation to

XX

PT isolate candidate molecules for rational drug discovery programs.

XX

PS Claim 14; SEQ ID NO 35434; 1766pp; English.

XX

XX

CC The invention relates to an isolated nucleic acid comprising any one of

XX

CC the 6213 antisense sequences given in the specification where expression

XX

CC of the nucleic acid inhibits proliferation of a cell. Also included are:

XX

CC (1) a vector comprising a promoter operably linked to the nucleic acid

XX

CC encoding a polypeptide whose expression is inhibited by the antisense

XX

CC nucleic acid; (2) a host cell containing the vector; (3) an isolated

XX

CC polypeptide or its fragment whose expression is inhibited by the

XX

CC antisense nucleic acid; (4) an antibody capable of specifically binding

XX

CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular

XX

CC proliferation or the activity of a gene in an operon required for

XX

CC proliferation; (7) identifying a compound that influences the activity of

XX

CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification. The sequence data for this patent did
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX

SQ Sequence 3211 BP; 1137 A; 504 C; 632 G; 938 T; 0 U; 0 Other;

Query Match 40.9%; Score 28.2; DB 8; Length 3211;

Best Local Similarity 64.6%; Pred. No. 4.2;

Matches 42; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

Qy 2 ATGCTAATCTAGGAGTTTCAATTTTCAGCAGGATATCAGACGGCTTTACTGTAATA 61

Db 593 ATGACAATATTGAAGAAGTTTCAATCAACAACAGCAGTATTAGTTATTACAGATGATG 652

Qy 62 TCACT 66

Db 653 TCATT 657

RESULT 3

ADR52731/C

ID ADR52731 standard; DNA; 127917 BP.

XX

AC ADR52731;

XX

DT 18-NOV-2004 (first entry)

XX

DE Drug therapy altered expressed gene #82.

XX

KW drug activity monitoring; expression profile; gene expression;

XX

KW peripheral blood sample; peripheral blood mononuclear cell; drug therapy;

XX

KW CCI-779; immunosuppressant; rapamycin; mammalian target of rapamycin;

XX

OS Homo sapiens.

XX

PN WO2004072265-A2.

XX

PD 26-AUG-2004.

XX

PF 11-FEB-2004; 2004WO-US004118.

XX

PR 11-FEB-2003; 2003US-0446133P.

XX

PR 03-APR-2003; 2003US-0459782P.

XX

PR 23-JAN-2004; 2004US-0538246P.

XX

PA (AMHP) WYETH.

XX

PA (BURC/) BURCZYNSKI M.

XX

PA (TWIN/) TWINE N.

XX

PA (DORN/) DORNER A J.

XX

PA (TREP/) TREPICCHIO W L.

XX

XX

PI Burczynski M, Twine N, Dornier AJ, Trepicchio WL;

XX

XX WPI; 2004-642301/62.

XX

PT Monitoring drug activities in vivo comprises comparing an expression
PT profile of a gene in a peripheral blood sample of a patient before and
PT after drug therapy.
XX
PS
PS
XX
XX
CC The invention relates to a method of monitoring drug activities in vivo
CC by comparing an expression profile of at least one gene in a peripheral
CC blood sample of a patient to a reference expression profile of the at
CC least one gene, where the at least one gene is differentially expressed
CC in peripheral blood mononuclear cells (PBMCs) of patients who have a non-
CC blood disease and are subjected to a drug therapy as compared to PBMCs
CC isolated from the patient before the drug therapy, and where the patient
CC has the non-blood disease and is being treated by the drug therapy. The
CC method, kit, and nucleic acid array are useful for monitoring drug
CC activities in vivo. The drug is especially CCI-779, an ester analogue of
CC the immunosuppressant rapamycin which is a potent inhibitor of the
CC mammalian target of rapamycin (mTOR). This sequence represents a gene
CC expressed in PBMC altered by the drug therapy. (Note: this sequence does
CC no form part of the printed specification but was obtained in electronic
CC format from WIPO at ftp.wipo.int/pub/published_pct_sequences/).

XX
SQ Sequence 127917 BP; 35803 A; 28668 C; 28463 G; 34983 T; 0 U; 0 Other;

Query Match 38.8%; Score 26.8; DB 13; Length 127917;
Best Local Similarity 64.5%; Pred. No. 33;
Matches 40; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

Qy 1 GATGCTAATCTAGGAGTTTCAATTTCCAGCAGGTTATCAGACGGCTTTTACTGGTAAT 60
Db 114718 GATGCAATCTAGCAGGACCTGAAATCCAGGAGGCTTCACTGCTGCTTCCAGCAAAAG 114659

Qy 61 AT 62
Db 114658 AT 114657

RESULT 4
AAC51660/c
ID AAC51660 standard; DNA; 1533 BP.
XX
AC AAC51660;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 69324.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway; metabolic pathway;
KW promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-00301439.
XX
XX 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 29-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.

PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135639P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.

PR 21-JUL-1999;	99US-0144814P.	PR 18-OCT-1999;	99US-0159584P.
PR 21-JUL-1999;	99US-0145086P.	PR 21-OCT-1999;	99US-0160741P.
PR 21-JUL-1999;	99US-0145088P.	PR 21-OCT-1999;	99US-0160767P.
PR 22-JUL-1999;	99US-0145088P.	PR 21-OCT-1999;	99US-0160768P.
PR 22-JUL-1999;	99US-0145088P.	PR 21-OCT-1999;	99US-0160770P.
PR 22-JUL-1999;	99US-0145089P.	PR 21-OCT-1999;	99US-0160814P.
PR 22-JUL-1999;	99US-0145192P.	PR 21-OCT-1999;	99US-0160815P.
PR 23-JUL-1999;	99US-0145145P.	PR 22-OCT-1999;	99US-0160980P.
PR 23-JUL-1999;	99US-0145218P.	PR 22-OCT-1999;	99US-0160981P.
PR 23-JUL-1999;	99US-0145224P.	PR 22-OCT-1999;	99US-0160989P.
PR 26-JUL-1999;	99US-0145276P.	PR 25-OCT-1999;	99US-0161404P.
PR 27-JUL-1999;	99US-0145913P.	PR 25-OCT-1999;	99US-0161405P.
PR 27-JUL-1999;	99US-0145918P.	PR 25-OCT-1999;	99US-0161406P.
PR 27-JUL-1999;	99US-0145919P.	PR 26-OCT-1999;	99US-0161359P.
PR 28-JUL-1999;	99US-0145951P.	PR 26-OCT-1999;	99US-0161360P.
PR 02-AUG-1999;	99US-0146386P.	PR 26-OCT-1999;	99US-0161361P.
PR 02-AUG-1999;	99US-0146388P.	PR 28-OCT-1999;	99US-0161920P.
PR 02-AUG-1999;	99US-0146389P.	PR 28-OCT-1999;	99US-0161992P.
PR 03-AUG-1999;	99US-0147038P.	PR 28-OCT-1999;	99US-0161993P.
PR 04-AUG-1999;	99US-0147204P.	PR 29-OCT-1999;	99US-0162142P.
PR 04-AUG-1999;	99US-0147302P.		
PR 05-AUG-1999;	99US-0147192P.	Query Match 38.3%; Score 26.4; DB 3; Length 1533;	
PR 05-AUG-1999;	99US-0147260P.	Best Local Similarity 65.0%; Pred. No.16;	
PR 06-AUG-1999;	99US-0147303P.	Matches 39; Conservative 0; Mismatches 21; Indels 0; Gaps 0;	
PR 06-AUG-1999;	99US-0147416P.		
PR 09-AUG-1999;	99US-0147493P.		
PR 09-AUG-1999;	99US-0147935P.		
PR 10-AUG-1999;	99US-0148171P.	Qy 3 TGCTAATACGTAGGAGCTTTCAATTTTCAGCAGGGGTATCAGAACGGCTTTTACTGGTAATAT 62	
PR 11-AUG-1999;	99US-0148319P.	Db 546 TGCTACACACTAGGAACAATAATCTTCAAAGGGTTCACAAACCTCGACAGATATAT 487	
PR 12-AUG-1999;	99US-0148341P.		
PR 13-AUG-1999;	99US-0148565P.		
PR 16-AUG-1999;	99US-0148684P.		
PR 17-AUG-1999;	99US-0149368P.		
PR 17-AUG-1999;	99US-0149175P.		
PR 18-AUG-1999;	99US-0149426P.		
PR 20-AUG-1999;	99US-0149722P.		
PR 20-AUG-1999;	99US-0149723P.		
PR 20-AUG-1999;	99US-0149929P.		
PR 23-AUG-1999;	99US-0149902P.		
PR 23-AUG-1999;	99US-0149930P.		
PR 25-AUG-1999;	99US-0150566P.		
PR 26-AUG-1999;	99US-0150884P.		
PR 27-AUG-1999;	99US-0151065P.		
PR 27-AUG-1999;	99US-0151066P.		
PR 27-AUG-1999;	99US-0151080P.		
PR 30-AUG-1999;	99US-0151303P.		
PR 31-AUG-1999;	99US-0151438P.		
PR 01-SEP-1999;	99US-0151930P.		
PR 07-SEP-1999;	99US-0152363P.		
PR 10-SEP-1999;	99US-0153070P.		
PR 13-SEP-1999;	99US-0153758P.		
PR 15-SEP-1999;	99US-0154018P.		
PR 16-SEP-1999;	99US-0154039P.		
PR 20-SEP-1999;	99US-0154779P.		
PR 22-SEP-1999;	99US-0155139P.		
PR 23-SEP-1999;	99US-0155488P.		
PR 24-SEP-1999;	99US-0155659P.		
PR 28-SEP-1999;	99US-0156458P.		
PR 29-SEP-1999;	99US-0156596P.		
PR 04-OCT-1999;	99US-0157117P.		
PR 05-OCT-1999;	99US-0157753P.		
PR 06-OCT-1999;	99US-0157865P.		
PR 07-OCT-1999;	99US-0158029P.		
PR 08-OCT-1999;	99US-0158232P.		
PR 12-OCT-1999;	99US-0158369P.		
PR 13-OCT-1999;	99US-0159293P.		
PR 13-OCT-1999;	99US-0159294P.		
PR 13-OCT-1999;	99US-0159295P.		
PR 14-OCT-1999;	99US-0159329P.		
PR 14-OCT-1999;	99US-0159330P.		
PR 14-OCT-1999;	99US-0159331P.		
PR 14-OCT-1999;	99US-0159637P.		
PR 14-OCT-1999;	99US-0159638P.		
AAC46785/C			
ID AAC46785 standard; DNA; 1707 BP.			
XX AAC46785;			
XX AC			
XX 18-OCT-2000 (first entry)			
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 51395.			
XX Hybridisation assay; genetic mapping; gene expression control;			
KW protein identification; signal transduction pathway; metabolic pathway;			
KW promoter; termination sequence; ss.			
XX Arabidopsis thaliana.			
OS Arabidopsis thaliana.			
XX EP1033405-A2.			
XX 06-SEP-2000.			
PF 25-FEB-2000; 2000EP-00301439.			
XX 25-FEB-1999;	99US-0121825P.		
PR 05-MAR-1999;	99US-0123180P.		
PR 09-MAR-1999;	99US-0123548P.		
PR 23-MAR-1999;	99US-0125788P.		
PR 25-MAR-1999;	99US-0126284P.		
PR 29-MAR-1999;	99US-0126785P.		
PR 01-APR-1999;	99US-0127462P.		
PR 06-APR-1999;	99US-0128234P.		
PR 08-APR-1999;	99US-0128714P.		
PR 16-APR-1999;	99US-0129845P.		
PR 19-APR-1999;	99US-0130077P.		
PR 21-APR-1999;	99US-0130449P.		
PR 23-APR-1999;	99US-0130510P.		
PR 28-APR-1999;	99US-0130891P.		
PR 30-APR-1999;	99US-0131449P.		
PR 30-APR-1999;	99US-0132048P.		
PR 30-APR-1999;	99US-0132407P.		
PR 04-MAY-1999;	99US-0132484P.		
PR 05-MAY-1999;	99US-0132485P.		
PR 06-MAY-1999;	99US-0132486P.		
PR 06-MAY-1999;	99US-0132487P.		

PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144684P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.

PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 28-SEP-1999; 99US-0155659P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159233P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.

```
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 38.3%; Score 26.4; DB 3; Length 1707;
Best Local Similarity 65.0%; Pred. No. 17;
Matches 39; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 3 TGCTAATCTAGGAGTTTCATTTTCAGCAGGGGTATCAGAACGGCTTTACTGGTAATAT 62
DB 591 TGCTAACACTCTAGGACAATATCTTCAAAAGGGTTCCACAAAACCTCGACAGATAATAT 532

RESULT 6
ADQ63861/c
ID ADQ63861 standard; cDNA; 5978 BP.
AC ADQ63861;
XX
DT 07-OCT-2004 (first entry)
XX
DE Novel human cDNA sequence #1022.
XX
KW ss; gene; osteopathic; neuroprotective; nootropic; antiparkinsonian;
KW cytostatic; gene therapy; diagnostic marker; morbid state; osteoporosis;
KW neurological disease; Alzheimer's disease; Parkinson's disease; dementia;
KW cancer.
XX
OS Homo sapiens.
XX
PN EP1440981-A2.
XX
PD 28-JUL-2004.
XX
PF 21-JAN-2004; 2004EP-00001196.
XX
PR 21-JAN-2003; 2003JP-00102206.
PR 09-MAY-2003; 2003JP-00131392.
XX
PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI Yamamoto J, Isono Y, Nagai K, Irie R;
XX
DR WPI; 2004-535376/52.
DR P-PSDB; ADQ66049.
XX
XX Novel 2495 cDNA, useful for treating osteoporosis, neurological diseases,
PT Alzheimer's diseases, Parkinson's diseases, dementia and various cancers.
XX
PS Claim 1; SEQ ID NO 1022; 2449pp; English.
XX
XX The invention relates to 2495 novel polynucleotides (I) and their encoded
CC polypeptides, sequences hybridizing to these nucleotides, sequences
CC encoding partial polypeptides and sequences having 70% or 90% identity to
CC the nucleotide and protein sequences. The nucleotides and polypeptides
CC are useful as diagnostic markers or therapeutic target for the diseases
CC or morbid states. They are also useful for treating osteoporosis,
CC neurological diseases, Alzheimer's diseases, Parkinson's diseases,
CC dementia and various cancers. This sequence corresponds to a nucleotide
XX sequence of the invention.
XX
XX Sequence 5978 BP; 1691 A; 1343 C; 1230 G; 1714 T; 0 U; 0 Other;
```

```
Query Match 38.0%; Score 26.2; DB 12; Length 5978;
Best Local Similarity 63.5%; Pred. No. 27;
Matches 40; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 4 GCTAATCTAGGAGTTTCATTTTCAGCAGGGGTATCAGAACGGCTTTACTGGTAATATC 63
DB 4304 GCTACAGTCTTAGAAATTCAATTTTCATGAGGAAGCTGGCAATGATTAAGTAATATC 4245

QY 64 ACT 66
DB 4244 ACT 4242

RESULT 7
ADS60721/c
ID ADS60721 standard; cDNA; 845 BP.
XX
AC ADS60721;
XX
DT 02-DEC-2004 (first entry)
XX
DE Bacterial polynucleotide #12708.
XX
KW Recombinant DNA construct; transformed plant; improved plant property;
KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
KW pathogen tolerance; pest tolerance; plant disease resistance;
KW cell cycle pathway modification; plant growth regulator;
KW homologous recombination; seed oil yield; protein yield; carbohydrate;
KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
KW bacterial polynucleotide; gene; ss.
XX
OS Bacteria.
XX
PN US2003233675-A1.
XX
PD 18-DEC-2003.
XX
PF 20-FEB-2003; 2003US-00369493.
XX
PR 21-FEB-2002; 2002US-0360039P.
XX
PA (CAOY/) CAO Y.
PA (HINK/) HINKLE G J.
PA (SLAT/) SLATER S C.
PA (CHEN/) CHEN X.
PA (GOLD/) GOLDMAN B S.
XX
PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
XX
DR WPI; 2004-061375/06.
XX
XX New recombinant DNA construct comprising a promoter positioned to provide
PT for expression of a polynucleotide encoding a polypeptide from a
PT microbial source, useful for producing plants with improved properties.
XX
PS Claim 1; SEQ ID NO 36395; 122pp; English.
XX
XX The invention relates to a recombinant DNA construct comprising a
CC promoter functional in a plant cell, where the promoter is positioned to
CC provide for expression of a polynucleotide encoding a polypeptide from a
CC microbial source. The invention also relates to a transformed plant
CC comprising the recombinant DNA construct and a method of producing a
CC transformed plant having an improved property. The plant is a crop plant
CC such as maize or soybean. The method of producing a transformed plant
CC having an improved property comprises transforming a plant with the
CC recombinant DNA construct and growing the transformed plant, where the
CC polynucleotide or polypeptide is useful for improving plant properties.
CC The recombinant DNA construct is useful for producing plants with
CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC increased resistance to plant disease, better growth rate by modification
CC of the cell cycle pathway with plant growth regulators, increased rate of
```



```
XX WPI; 1998-240529/21.
DR P-PSDB; AAW55974, AAW55975.
XX
XX Nucleic acids encoding Mi polypeptide(s) conferring nematode resistance -
PT useful to produce transgenic plants resistant to these and other pests,
PT and in marker-aided selection to assess cultivars for resistance.
XX
XX Claim 11; Page 24-39; 55pp; English.
XX
XX This is the nucleotide sequence of the tomato Mi locus associated with
CC nematode resistance, obtained from a bacterial artificial chromosome
CC (BAC). Mi was localised by genetic analysis to a region of the tomato
CC genome of about 65 kb. DNA corresponding to this region was cloned into
CC BAC vectors. Sequence analysis of a 52 kb BAC3 insert identified 3 open
CC reading frames, one of which is probably a pseudogene. By RNA blot
CC analysis, transcripts of approximately 4 kb corresponding to copy 1 and
CC copy 2 were found in both resistant and susceptible tomato roots and in
CC leaves of resistant but not susceptible plants. cDNA sequences
CC corresponding to full-length transcripts of copy 1 (see. AAV26082) and
CC copy 2 (see AAV26083) were obtained. The encoded polypeptides (see
CC AAW55974-75) are 91% identical and contain structural features similar to
CC known plant resistance genes (R genes) of the nucleotide binding site/
CC leucine-rich repeat (NBS/LRR) family. A recombinant expression cassette
CC comprising an Mi polynucleotide and an operably linked plant promoter can
CC be used to enhance nematode resistance in plants especially tomatoes
CC (claimed). Transgenic plants can also be constructed using a Mi promoter
CC with heterologous genes; the Mi promoters can be used to express a
CC variety of genes in the same temporal and spatial patterns and at similar
CC levels to resistance genes
XX
XX Sequence 51952 BP; 17728 A; 8444 C; 8647 G; 17111 T; 0 U; 22 Other;
SQ
Query Match 36.8%; Score 25.4; DB 2; Length 51952;
Best Local Similarity 64.4%; Pred. No. 87;
Matches 38; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
Qy 1 GATGCTAATCTGTTAGGAGTTTCAATTTTCAGCAGGGTATCAGACGGCTTTACTGGTAA 59
Db 6714 GGTGATAGTAATCTCTGGAGTTGTCAGGGTCAAAAGGGTATCAGCAGAACTTGATTGAAA 6772
RESULT 10
ABD33333
ID ABD333333 standard; DNA; 66681 BP.
XX
XX ABD333333;
AC
XX
XX 18-NOV-2004 (first entry)
DT
XX
XX Murine cancer-associated (CA) gene MD07-060.
DE
XX
XX Mouse; cancer-associated protein; CAP; cancer-associated gene; CA; gene;
KW ds; cancer; cytostatic.
XX
XX Mus musculus.
OS
XX
XX WO2004058146-A2.
FN
XX
XX 15-JUL-2004.
PD
XX
XX 15-DEC-2003; 2003WO-US040081.
PF
XX
XX 17-DEC-2002; 2002US-00322281.
PR
XX
XX (SAGR-) SAGRES DISCOVERY INC.
PA
XX
XX Morris DW, Melandro MS;
PI
XX
XX WPI; 2004-499109/47.
DR
XX
XX Novel human cancer associated protein encoded within open reading frame
PT of cancer associated gene, useful as targets for diagnosing cancer.
PT
```

```
XX Disclosure; SEQ ID NO 411; 182pp; English.
PS
XX
XX The invention relates to cancer-associated proteins (CAP) and the cancer-
CC associated (CA) nucleic acids encoding them. The invention also relates
CC to a method for treating cancers involving administering to a patient an
CC inhibitor of CAP, and a method of screening for anticancer activity in a
CC potential drug involving providing a cell that expresses a CA gene,
CC contacting a tissue sample derived from a cancer cell with an anticancer
CC drug candidate and monitoring the effect of the anticancer drug candidate
CC on expression of the CA gene. The CAP proteins are useful for detecting
CC cancer associated with expression of a CAP protein in a test cell sample
CC and for screening for a bioactive agent capable of modulating the
CC activity of a CAP protein. The CA nucleic acids are useful for diagnosing
CC cancer, involving determining the expression of a CA nucleic acid in a
CC tissue. This sequence represents a murine CA gene of the invention. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 66681 BP; 15710 A; 15622 C; 17256 G; 18047 T; 0 U; 46 Other;
SQ
Query Match 36.8%; Score 25.4; DB 13; Length 66681;
Best Local Similarity 64.4%; Pred. No. 92;
Matches 38; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
Qy 6 TAATACCTGTAGGAGTTTCAATTTTCAGCAGGGTATCAGACGGCTTTACTGGTATATCA 64
Db 7130 TGATTTTGTAGGAGTTTCTGGGTCAAGTATAGTAGCAGCCATCTTCATAGGTCACCTCA 7188
RESULT 11
ABN21249
ID ABN21249 standard; cDNA; 322 BP.
XX
XX ABN21249;
AC
XX
XX 24-JUN-2002 (first entry)
DT
XX
XX Human ORFX polynucleotide sequence SEQ ID NO:10975.
DE
XX
XX Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
KW hypertension; hypothyroidism; cholesterol ester storage disease;
KW immune deficiency; immune disorder; infectious disease;
KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
KW myasthenia gravis; gene; ss.
XX
XX Homo sapiens.
OS
XX
XX WO200192523-A2.
FN
XX
XX 06-DEC-2001.
PD
XX
XX 29-MAY-2001; 2001WO-US010836.
PF
XX
XX 30-MAY-2000; 2000US-0206132P.
PR
XX
XX 29-AUG-2000; 2000US-0228716P.
PR
XX
XX (CURA-) CURAGEN CORP.
PA
XX
XX Shimkets RA, Leach MD;
PI
XX
XX WPI; 2002-106308/14.
DR
XX
XX P-PSDB; ABP05497.
DR
XX
XX Novel human polypeptides and polynucleotides useful for diagnosing,
PT preventing and treating cardiovascular disease, neurodegenerative,
PT hyperproliferative disorders and autoimmune disorders.
XX
XX Disclosure; SEQ ID NO 10975; 1037pp; English.
PS
```


CC such as pig selected for having desired genotypic or potential phenotypic
CC properties; (4) a transgenic animal comprising N1 or N2; and (5) sperm or
CC an embryo derived from the animal of (3) or (4). N1 or its fragment is
CC useful for selecting an animal destined for slaughter or a breeding
CC animal having desired genotypic or potential phenotypic properties. The
CC properties are related to muscle mass and/or fat deposition. The sperm or
CC an embryo are useful in breeding animals destined for slaughter. AA65418
CC to AA65524 represent contigs 1 to 10 and 19 to 115 which were isolated
CC from porcine BAC-P1GF2-1 which contains the INS and IGF2 (insulin-like
CC growth factor) genes. These sequences were used in an example from the
CC present invention for generating a reference sequence of IGF2 and
CC flanking loci in the pig
XX
SQ Sequence 500 BP; 163 A; 115 C; 132 G; 90 T; 0 U; 0 Other;

Query Match 36.2%; Score 25; DB 3; Length 500;
Best Local Similarity 61.5%; Pred. No. 42;
Matches 40; Conservative 0; Mismatches 25; Indels 0; Gaps 0;
QY 1 GATGCTAATCTGTAGGAGTTTCATTTTCAGCAGGGTATCAGACGGCTTACTGGTAAT 60
DB 429 GATTCATGTCTGTATGTGTGTTGCAAGTGATCACAGCAGCGGTGCGGTTAATTCATT 370
QY 61 ATCAC 65
DB 369 TCCAC 365

RESULT 14
ADQ24808
ID ADQ24808 standard; DNA; 596 BP.
XX
AC ADQ24808;
XX
DT 26-AUG-2004 (first entry)
XX
DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 7628.
XX
KW soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
KW ds.
XX
OS Homo sapiens.
XX
FN WO2004048938-A2.
XX
PD 10-JUN-2004.
XX
PF 26-NOV-2003; 2003WO-US038193.
XX
PR 26-NOV-2002; 2002US-0429739P.
XX
PA (PROT-) PROTEIN DESIGN LABS INC.
XX
PI Aziz N, Ginsburg WM, Zlotnik A;
XX
DR WPI; 2004-441208/41.
XX
PT Early detection of soft tissue sarcoma comprises determining expression
PT of a gene in a first soft tissue sample and a normal soft tissue sample
PT and comparing the gene expression, also useful in treating soft tissue
PT sarcoma.

XX
XX Example 2; SEQ ID NO 7628; 210pp; English.
XX
CC The invention relates to a novel method for detecting soft tissue sarcoma
CC which comprises obtaining a first soft tissue sample from an individual
CC and a normal soft tissue sample from the same or different individual,
CC determining the expression of a gene in both samples and comparing the
CC expression of the gene in both soft tissue samples, where a higher level
CC of protein expression in the first soft tissue sample indicates the
CC presence of soft tissue sarcoma. The method of the invention has
CC cytostatic applications and may be useful for detecting soft tissue
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic

CC acid sequences may be useful in diagnostic and screening applications.
CC The current sequence is that of a human soft tissue sarcoma-upregulated
CC DNA of the invention. The current sequence is not shown within the
CC specification per se but was submitted in CD format by the inventor.

XX
SQ Sequence 596 BP; 158 A; 112 C; 105 G; 221 T; 0 U; 0 Other;
Query Match 36.2%; Score 25; DB 12; Length 596;
Best Local Similarity 64.9%; Pred. No. 43;
Matches 37; Conservative 0; Mismatches 20; Indels 0; Gaps 0;
QY 12 TGTAGGAGTTTCATTTTCAGCAGGGTATCAGACGGCTTACTGGTAATCACTAC 68
DB 102 TTTAGGTTTTCTTTTAAATGAGGCTCTACAACTATTTTCCTTTTATTACCACTAC 158

RESULT 15
ACCA46307/C
ID ACC46307 standard; cDNA; 1308 BP.

XX
AC ACC46307;
XX
DT 02-JUN-2003 (first entry)
XX
DE Human dithp transcription factor-encoding cDNA.
XX
KW Human; dithp; diagnostic and therapeutic polynucleotide; diagnosis;
KW cancer; cell proliferative disorder; autoimmune disorder;
KW inflammatory disorder; infection; hormonal disorder; metabolic disorder;
KW neurological disorder; gastrointestinal disorder; transport disorder;
KW connective tissue disorder; drug screening; proteome analysis;
KW gene therapy; antisense therapy; genotyping; transgenic animal; knock in;
KW disease model; toxicological testing; transcript imaging;
KW transcription factor; gene; ss.

OS Homo sapiens.
XX
PN WO200297031-A2.
XX
PD 05-DEC-2002.
XX
PF 27-MAR-2002; 2002WO-US010056.
XX
PR 28-MAR-2001; 2001US-0279619P.
PR 29-MAR-2001; 2001US-0280067P.
PR 29-MAR-2001; 2001US-0280068P.
PR 16-MAY-2001; 2001US-0291280P.
PR 17-MAY-2001; 2001US-0291829P.
PR 17-MAY-2001; 2001US-0291849P.
PR 19-JUN-2001; 2001US-0299428P.
PR 20-JUN-2001; 2001US-0299776P.
PR 20-JUN-2001; 2001US-0300001P.

XX (INCY-) INCYTE GENOMICS INC.
XX
PI Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;
PI Dufour GE, Hillman JL, Yu JY, Tuason O, Yap PE, Ameshey SR;
PI Daughtery SC, Dam TC, Liu TF, Nguyen DA, Kleeefeld Y, Gerstin EH;
PI Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B;
PI Flores V, Marwaha R, Lo A, Lan RY, Urashka ME;
XX
DR WPI; 2003-129518/12.
DR P-FSDB; ABR41367.

XX
PT Novel human diagnostic and therapeutic polypeptide useful for identifying
PT test compound which specifically binds to a polypeptide encoded by human
PT diagnostic and therapeutic polynucleotide, and to induce antibodies.
XX
PS Claim 2; SEQ ID NO 228; 591pp; English.

XX
CC The invention relates to novel human diagnostic and therapeutic
CC polynucleotides designated dithp (ACC46080-ACC46749) and to their encoded
CC proteins (DITHP; ABR41136-ABR41812). The invention also relates to

Search completed: November 20, 2005, 07:07:16
Job time : 286 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 2005, 06:10:24 ; Search time 1675 Seconds
(without alignments)
1568.021 Million cell updates/sec

Title: US-10-780-250-4
Perfect score: 69
Sequence: 1 gatgctaatacttagtagt.....ttactggtaatactactaca 69

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues
Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:
1: gb_est1:
2: gb_est2:
3: gb_hc:
4: gb_est3:
5: gb_est4:
6: gb_est5:
7: gb_est6:
8: gb_gsa1:
9: gb_gsa2:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	28.4	41.2	883	5	BX702860
C 2	28.4	41.2	1040	9	CL084804
C 3	28	40.6	701	8	CC317613
C 4	27.8	40.3	545	8	AQ570246
C 5	27.2	39.4	289	1	AA325500
C 6	27.2	39.4	454	1	AV768149
C 7	27.2	39.4	467	4	BF997598
C 8	27.2	39.4	503	5	BA451647
C 9	27.2	39.4	610	5	BU672369
C 10	27.2	39.4	909	9	CNS061X2
C 11	27	39.1	446	8	AQ543352
C 12	27	39.1	493	9	CL900277
C 13	27	39.1	723	8	AZ243876
C 14	27	39.1	916	9	CL477612
C 15	26.8	38.8	701	8	BH998573
C 16	26.6	38.6	672	7	CO239569
C 17	26.6	38.6	753	7	CNS833609
C 18	26.4	38.3	667	8	BZ841925
C 19	26.4	38.3	1683	3	CNS0A2J9
C 20	26.4	38.3	1782	3	CNS0A2EW
C 21	26.2	38.0	501	8	AQ626976
C 22	26.2	38.0	685	8	BZ217305
C 23	26	37.7	507	7	CO906945
C 24	26	37.7	576	4	BM686321

25	26	37.7	617	9	CE614541
C 26	26	37.7	649	8	AZ890160
C 27	26	37.7	657	6	CA057268
C 28	26	37.7	711	7	CN187254
C 29	26	37.7	767	6	CD576567
C 30	26	37.7	769	6	CB512076
C 31	26	37.7	990	8	CC194365
C 32	26	37.7	1135	9	CL147131
C 33	25.8	37.4	432	8	AQ651037
C 34	25.8	37.4	548	4	BG625709
C 35	25.8	37.4	549	8	AQ951889
C 36	25.8	37.4	557	9	TA392804P
C 37	25.8	37.4	577	8	AQ656315
C 38	25.8	37.4	585	5	BW352543
C 39	25.8	37.4	624	5	BW350754
C 40	25.8	37.4	624	8	AZ220668
C 41	25.8	37.4	627	8	AQ654644
C 42	25.8	37.4	638	5	BW361520
C 43	25.8	37.4	651	8	AZ982438
C 44	25.8	37.4	694	5	BW265835
C 45	25.8	37.4	712	7	CF360068

ALIGNMENTS

RESULT 1
BX702860/c
LOCUS BX702860 XGC-tadpole Xenopus tropicalis cDNA clone TTPA022d21 3',
DEFINITION mRNA sequence.
ACCESSION BX702860
VERSION BX702860.1 GI:38365067
KEYWORDS EST.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
Xenopodinae; Xenopus; Silurana.
REFERENCE 1 (bases 1 to 883)
Croning,M.D.R., Ashurst,J.L., Taylor,R., Zorn,A.M. and Rogers,J.
Sanger Xenopus tropicalis EST project 2001 (11_2003)
TITLE Unpublished (2003)
JOURNAL
COMMENT Contact: Croning MDR
Sanger Institute
Hinnton, Cambridgeshire, CB10 1SA, UK
Email: tropesanger.ac.uk
Sanger Xenopus tropicalis EST project 2001
TROPICALIS_SEQUENCE_ID: TTPA022d21.q1kT7
Sequencing primer: T7
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Nigel Garrett.
cDNA was oligo dt primed from Sug of poly A+ RNA from tadpole
embryos. EcoRI-NotI cut cDNA was then ligated into pCS107 with
EcoRI at the 5' end and NotI at the 3' end.
Vector: pCS107; Site 1: EcoRI; Site 2: NotI
Host: Escherichia coli DH10B.
Location/Qualifiers
1..883
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clones="TTPA022d21"
/dev_stage="tadpole (stage 35-40)"
/lab_host="E. coli DH10B"
/clone_lib="XGC-tadpole"
/note="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA
was oligo dt primed from Sug of poly A+ RNA from tadpole
embryos. EcoRI-NotI cut cDNA was then ligated into pCS107
with EcoRI at the 5' end and NotI at the 3' end"

ORIGIN

Query Match 41.2%; Score 28.4; DB 5; Length 883;

Best Local Similarity	66.1%;	Pred. NO. 25;	
Matches	41; Conservative	0; Mismatches	21; Indels
			0; Gaps

	GCTAATACTGTAGGAGTTTCAATTTCAGCAGGGTATCAGAACCGCTTTACTGGTAATATC	63
Qy		
Db	594 GATAGTGTTTAGAAAGTTAGAATGTAAACGTAGTAGTAGAGGGGCCCTTAAGGGTAAATAGC	535
Qy	64 AC 65	
Db	534 AC 533	

RESULT 2				
CL084804				
LOCUS	CL084804	1040 bp	DNA	linear
DEFINITION	ISB1-3P20_Sp6.1 ISB1 <i>Xenopus tropicalis</i> genomic clone ISB1-3P20, genomic survey sequence.			

ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
CL084804				
CL084804.1	GI:40559709			
GSS				
xenopus tropicalis				western clawed frog
xenopus tropicalis				
Eukaryota; Metazoa;				Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura;				Mesobatrachia; Pipidoidea; Pipidae;
Xenopodinae; Xenopus;				Silurana.

REFERENCE
1 (bases 1 to 1040)
AUTHORS Kremitzki, C., Carter, J., McPherson, J., Warren, W., Graves, T.,
Mardis, E. and Wilson, R.
TITLE A physical map of the xenopus tropicalis genome
JOURNAL Unpublished (2003)
COMMENT Contact: Richard K Wilson

Genome Sequencing Center
Washington University School of Medicine
Email: submissions@wustl.edu
Insert Length: 75000 Std Error: 0.00
Seq primer: Sp6 ATTAGGTGACACTATAG
Class: BAC ends
High quality sequence start: 42
High quality sequence stop: 768.

```

FEATURES
    source
        Location/Qualifiers
            1..1040
                /organism="Xenopus tropicalis"
                /mol_type="genomic DNA"
                /db_xref="taxon:8364"
                /clone="ISB1-3P20"
                /clone_lib="ISB1"
                /note="Vector: pBelOBAC11; ISB-1 Xenopus tropicalis BAC
                Library Segment 1"

```

ORIGIN	Library Segment 1			
Query Match	41.2%	Score 28.4;	DB 9;	Length 1040;
Best Local Similarity	66.1%;	Pred. No. 26;		
Matches	41;	Conservative	21;	Indels 0;
		Mismatches	21;	Gaps 0;

QY 4 GCTAATACTGTAGGAGTTTCAATTTTCAGCAGGGTATCAGAACGGCTTTACTGGTAATATC 63
127 COMBUSTION 0, HIGHRECTION 21, INCLIN 0, CAPS 0

D8 573 GATAGTTGTTTAGAAGTTAGAACTGTAACCTCAGTAGTAGAGGGGCCCTTAAGGGTAATATGC 632

Qy 64 AC 65
||
Db 633 AC 634

RESULT 3	CC317613	701 bp	DNA	linear	GSS 14-MAY-2003
LOCUS	CC317613/c				
DEFINITION	TAM32-25K5 Sp6.1 TAM32 Gallus gallus genomic clone TAM32-25K5,				
	genomic survey sequence.				
ACCESSION	CC317613				
VERSION	CC317613.1				
KEYWORDS	GSS.				
SOURCE	CC317613.1	GI:30711671			
ORGANISM	Gallus gallus (chicken)				
	Gallus gallus				
	Gallus gallus				

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 701)
Kremetzki,C., Higginbotham,J., Wylie,K., Carter,J., McPherson,J.,
Warren,W., Graves,T., Mardis,E. and Willson,R.
Gallus gallus BAC End Reads
Unpublished (2003)
Contact: Richard K. Willson

Genome Sequencing Center
Washington University School of Medicine
Email: submissions@wustl.wustl.edu
Insert Length: 152000 Std Error: 0.00
Seq primer: Sp6 ATTAGGTGACACTATAG
Class: BAC ends
High quality sequence start: 7
High quality sequence stop: 349.

FEATURES	source	Location/Qualifiers
		1. .701
		/organism="Gallus gallus"
		/mol_type="genomic DNA"
		/strain="Red Jungle Fowl"
		/db_xref="taxon:9031"
		/clone="TAM32-25K5"
		/sex="female"
		/cell_line="UCD001, inbred 256"
		/clone_lib="TAM32"
		/note=vector: pECBAC1; Site_1: EcoRI; Site_2: EcoRI;
		TAM32 Female Chicken library - for library and clone
		ordering information: http://www.hbz.tamu.edu

ORIGIN

Query Match 40.6%; Score 28; DB 8; Length 701;
Best Local Similarity 63.2%; Pred. No. 34;
Matches 43; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

Ordering information: <http://www.hb2.cam.ac.uk>

Qy 1 GATGCTAAATCTGTAGGAGTTCCTCAATTTTCAGCAGGGTATCAGAACGCCCTTTACTGGTAAT 60
330 GCTACTAAACTTTGGCATTCTGATTTTCAGCAGGGACTCAGAGCTCCTTAAGGAATAT 271

Qy	61 ATCACTAC 68
Db	270 ATCAGTTC 263

RESULT 4	AQ570246	545 bp	DNA	linear	GSS 01-JUN-1999
LOCUS	AQ570246/c				
DEFINITION	HS 5356 Al_F09 T7A RPCR-11 Human Male BAC Library Homo sapiens genomic clone Plates=932 Col=17 Row=K, genomic survey sequence.				
ACCESSION	AQ570246				
VERSION	AQ570246.1				
KEYWORDS	GSS.				
SOURCE	Homo sapiens (human)				

SOURCE: Homo sapiens (human);
ORGANISM: Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidei; Homo.
1 (bases 1 to 545)
Mahaïras, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T.,
Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D., and
Hood, L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
PUBMED 10449764
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu

Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering/bac.htm>) or from Research Genetics (<http://www.hsc.washington.edu>). BAC end Web Server: <http://www.hsc.washington.edu>

Seq primer: 17

Class: BAC ends

High quality sequence stop: 545.

Location/Qualifiers

1. 545

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/clone="Plate=932 Col=17 Row=K"

/sex="male"

/clone.lib="RPCI-11 Human Male BAC Library"

/note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACE3.6 vector at EcoRI sites"

ORIGIN

Query Match 40.3%; Score 27.8; DB 8; Length 545;

Best Local Similarity 67.9%; Pred. No. 38;

Matches 38; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Qy 14 TAGAGTTTCAATTCACAGGGTATCAGACGGCTTACTGGTAATATCACTACA 69

Db 92 TATGTGTTTGAATTTAGTTAGGTATAGACNAATTAACCTAATGAGATACCCACA 37

RESULT 5

AA325500/c

LOCUS EST28526 Cerebellum II Homo sapiens cDNA 5' end similar to EST

DEFINITION containing Alu repeat, mRNA sequence.

ACCESSION AA325500

VERSION AA325500.1 GI:1977765

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 289)

Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,

Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,

White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,

Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,

Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,

Glocke,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr.,

Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,

Morono-Palancas,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,

Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,

Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,

Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,

Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,

He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,

Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meisner,P.S., Olsen,H.,

Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,

Dillon,P.J., Fannon,M.R., Rosen,C.A., Hasseltine,W.A., Fields,C.,

Fraser,C.M. and Venter,J.C.

Initial assessment of human gene diversity and expression patterns

based upon 83 million nucleotides of cDNA sequence

Nature 377 (6547 Suppl), 3-174 (1995)

96026280

7566098

Contact: Kerlavage, AR

Bioinformatics

The Institute for Genomic Research

9712 Medical Center Drive, Rockville, MD 20850 USA

Tel: 3018699056

Fax: 3018699423

Email: arkerlav@tigr.org

For clone availability, additional sequence and expression

information related to this EST, please check the TIGR Human Gene

Index (<http://www.tigr.org/tdb/hgi/hgi.html>)

Seq primer: M13 Reverse.

Location/Qualifiers

1. 289

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="ATCC (inhost):126027"

/db_xref="taxon:9606"

/tissue_type="cerebellum"

/dev stage="adult"

/clone.lib="Cerebellum II"

/note="Organ: brain; Vector: pBluescript SK-; Site_1:

EcoRI; Site_2: XhoI"

ORIGIN

Query Match 39.4%; Score 27.2; DB 1; Length 289;

Best Local Similarity 72.9%; Pred. No. 57;

Matches 35; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 7 AATACTGTAGAGTTTCAATTTTCAGACGGGTATCAGAACGGCTTTACT 54

Db 114 AATTCGTATCAGTTTTCACCTTAAGCAGGGCATTAAGGACTCCTAGACT 67

RESULT 6

AV768149/c

LOCUS AV768149

DEFINITION AV768149 Lotus japonicus Young plants (two-weeks old) Lotus

corniculatus var. japonicus cDNA clone MIM223d11_f 3', mRNA

sequence.

ACCESSION AV768149

VERSION AV768149.1 GI:45351423

KEYWORDS EST.

SOURCE LOTUS

ORGANISM

Lotus corniculatus var. japonicus (Lotus japonicus)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;

Lotus.

1 (bases 1 to 454)

Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.

Characteristics of the Lotus japonicus Gene Repertoire Deduced from

Large-Scale Expressed Sequence Tag (EST) Analysis

Plant Mol. Biol. 54 (3), 405-414 (2004)

Contact: Erika Asamizu

The First Laboratory for Plant Gene Research

Kazusa DNA Research Institute

Yana 1532-3, Kisarazu, Chiba 292-0812, Japan

Email: asamizu@kazusa.or.jp, URL: <http://www.kazusa.or.jp/en/plant/>.

Location/Qualifiers

1. 454

/organism="Lotus corniculatus var. japonicus"

/mol_type="mRNA"

/variety="japonicus"

/db_xref="taxon:34305"

/clone="MIM223d11_f"

/tissue_type="Young plants (two-weeks old)"

/clone.lib="Lotus japonicus Young plants (two-weeks old)"

/notes="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:

XhoI; isolate=Miyakojima MG-20"

ORIGIN

Query Match 39.4%; Score 27.2; DB 1; Length 454;

Best Local Similarity 67.9%; Pred. No. 61;

Matches 38; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Qy 11 CTGTAGAGTTTCAATTTTCAGACGGGTATCAGAACGGCTTTACTGGTAATATCACT 66

Tel: 5105959773
Fax: 5105959818

Email: oanderan@pw.usda.gov

Sequences have been trimmed to remove vector sequence and low quality sequence with phred score less than 20

Seq primer: SK primer:

FEATURES

source
1..610
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="WHE3303_G08_N15"
/tissue_type="root"
/dev_stage="Full tillering stage"
/lab_host="E. coli SOLR"
/clone_lib="Chinese Spring wheat drought stressed root cDNA library"
/note="Vector: Lambda Uni-ZAP XR, excised phagemid; Site 1: EcoRI; Site 2: XhoI; Plants were grown under normal conditions, then drought stressed to 80%, 70% and 60% RWC at Texas Tech University (D. Zhang in HT Nguyen lab). Total RNA was prepared separately for roots collected at the three different drought conditions. Equal amounts of total RNA were pooled from all three samples, poly(A) RNA were purified, one cDNA library was made, and the cDNA clones were in vivo excised to give pBluescript SK(-) phagemids in the TJ Close lab at the University of California, Riverside (Fenton, Turuspekov). Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."

ORIGIN

Query Match 39.4%; Score 27.2; DB 5; Length 610;
Best Local Similarity 67.9%; Pred. No. 64;
Matches 38; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Oy 9 TACTGTAGGAGTTCAATTCAGCAGGGGTATCAGAACGGCTTTACTGTGTAATATCA 64
Db 550 TTCTCCAGCAVTTGCACTTCAGCAGGGGACCAAGATGTCATGCAATGAGTAATATCA 495

RESULT 10

CNS06UX2/c
LOCUS
DEFINITION
T3 end of clone AX0AA014D01 of library AX0AA from strain CBS 7064 of Pichia farinosa, genomic survey sequence.
ACCESSION
AL416412
VERSION
AL416412.1 GI:12196816
KEYWORDS
GSS.
SOURCE
Pichia farinosa
ORGANISM
Pichia farinosa
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Pichia.

REFERENCE

AUTHORS
Souciet, J.L., Aigle, M., Artiguenave, F., Blandin, G., Bolotin-Fukuhara, M., Bon, E., Brottier, P., Casaregola, S., de-Montigny, J., Dujon, B., Durrens, P., Lepingle, A., Llorente, B., Malpertuy, A., Neuvéglise, C., Ozier-Kalogeropoulos, O., Potier, S., Saurin, W., Tekala, F., Toffano-Nioche, C., Wesolowski-Louvel, M., Wincker, P. and Weissenbach, J.
TITLE
Yeast species for molecular evolution studies

Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies
FEBS Lett. 487 (1), 3-12 (2000)

JOURNAL
MEDLINE
20584711

PUBMED
11152876

REFERENCE

AUTHORS
de Montigny, J., Spohnner, C., Souciet, J., Tekala, F., Dujon, B., Wincker, P., Artiguenave, F. and Potier, S.
TITLE
Genomic exploration of the hemiascomycetous yeasts: 15. Pichia sorbitophila

JOURNAL
FEBS Lett. 487 (1), 87-90 (2000)
MEDLINE
20584725

PUBMED

REFERENCE
11152890
3 (bases 1 to 909)
Genoscope.
Direct Submission
Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail: seqref@genoscope.cns.fr - Web: www.genoscope.cns.fr)
JOURNAL
This GSS is part of a random genomic sequencing program of thirteen yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces exiguus, Saccharomyces servazii, Zygosaccharomyces rouxii, Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila, Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.

FEATURES

source
1..909
/organism="Pichia farinosa"
/mol_type="genomic DNA"
/strain="CBS 7064"
/db_xref="taxon:4920"
/clone="AX0AA014D01"
/clone_lib="AX0AA"
/note="end : T3"

ORIGIN

Query Match 39.4%; Score 27.2; DB 9; Length 909;
Best Local Similarity 72.9%; Pred. No. 68;
Matches 35; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
Oy 21 TTCAATTTTCAGCAGGGTATCAGAACGGCTTTACTGTGTAATATCTACTAC 68
Db 731 TTCAATATCAGCAGGATCTCAATACGACACACTGGTATATCTCTAC 684

RESULT 11

AQ543352
LOCUS
DEFINITION
RPC1-11-36018-TV RPC1-11 Homo sapiens genomic clone RPC1-11-36018, genomic survey sequence.

ACCESSION
AQ543352

VERSION
AQ543352.1 GI:4873636

KEYWORDS
GSS.

SOURCE
Homo sapiens (human)

ORGANISM
Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Zhao, S., Adams, M.D., Nierman, W., Malek, J., de Jong, P. and Venter, J.C.

REFERENCE
1 (bases 1 to 446)

Use of BAC End Sequences from Library RPC1-11 for Sequence-Ready

AUTHORS
Map Building

Unpublished (1997)

Other GSSs: RPC1-11-36018.TJ

Contact: Shaving Zhao, William Nierman, Mark Adams

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850

Tel: 301 838 0200

Fax: 301 838 0208

Email: hbe@tigr.org

Clones are derived from the human BAC library RPC1-11. For BAC

library availability, please contact Pieter de Jong

(pieter@dejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from

Research Genet cs (info@resgen.com). BAC end search page:

http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.

Seq primer: T7

Class: BAC ends.

Location/Qualifiers

1..446

```

/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="GDB:7638055"
/db_xref="taxon:9606"
/clone="RPCI-11-36018"
/sex="Male"
/cell_type="Lymphocytes"
/clone_lib="RPCI-11"
/note="Vector: pBACe3.6; Site 1: EcoRI; Site 2: EcoRI;
RPCI11 Human Male BAC Library"

ORIGIN
Query Match          39.1%; Score 27; DB 8; Length 446;
Best Local Similarity 66.1%; Pred. No. 72;
Matches 39; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

Qy 1 GATGCTAATCTAGGAGTTTCATTTTCACGAGGATATCAGACGGCTTTACTGGTAA 59
|||||
Db 27 GTTGCCATCTCTGTAGTTGATTCAGTTTCACGAGAGAGAGGGGCTCTACAGTAA 85
|||||

RESULT 12
CL900277/c
LOCUS
DEFINITION
abg59g12.y1 Soybean random, unfiltered genomic library glycine max
genomic, genomic survey sequence.
ACCESSION
CL900277
VERSION
CL900277.1 GI:51660910
KEYWORDS
GSS.
SOURCE
Glycine max (soybean)
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.
REFERENCE
1 (bases 1 to 493)
Nunberg,A., Bedell,J.A., Citek,R.W., Robbins,D., McMenamy,J.,
Peterson,S., Jones,S., Fries,S., Budiman,M.A., Nguyen,H. and
Stacey,G.
Methylation filtered genomic sequences from Glycine max
Unpublished (2004)
Contact: Gary Stacey
University of Missouri
108 Waters Hall, Columbia, MO 65211, USA
Tel: 573-884-1267
Fax: 573-882-0588
Email: stacey@missouri.edu
LidID: 230
Class: shotgun.
FEATURES
Location/Qualifiers
source
1..493
/organism="Glycine max"
/mol_type="genomic DNA"
/cultivar="Williams 82"
/db_xref="taxon:3947"
/tissue_type="Young leaves"
/clone_lib="Soybean random, unfiltered genomic library"
/note="Vector: pOT2; Site 1: BstXI; Randomly sheared
genomic DNA ranging from 0.7-1.5 kb were end repaired and
ligated to BstXI linkers prior to cloning in BstXI-cut
pOT2. LibID: 230"

ORIGIN
Query Match          39.1%; Score 27; DB 9; Length 493;
Best Local Similarity 62.7%; Pred. No. 73;
Matches 42; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

Qy 1 GATGCTAATCTAGGAGTTTCATTTTCACGAGGATATCAGACGGCTTTACTGGTAA 60
|||||
Db 282 GATGTCGGGCTGCAGAGCTTCTCTTTGTGCAATTGAATGGATCTGCTGCTGGGAAG 223
|||||
Qy 61 ATCACTA 67
|||||

```

```

Db 222 CACATTA 216

RESULT 13
AZ243876/c
LOCUS
DEFINITION
RPCI-23-88K11.TJ RPCI-23 Mus musculus genomic clone RPCI-23-88K11,
genomic survey sequence.
ACCESSION
AZ243876
VERSION
AZ243876.1 GI:8557067
KEYWORDS
GSS.
SOURCE
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE
1 (bases 1 to 723)
AUTHORS
Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S.,
Akinret,B., Levins,M., McGann,S., Tsengye,G., Geer,K., Krol,M., de
Jong,P. and Fraser,C.M.
Mouse BAC End Sequences from Library RPCI-23
Unpublished (1999)
Other GSSs: RPCI-23-88K11.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pieter@jeng.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
or from Resea ch Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/tdb/bac ends/mouse/bac_end_intro.html
Plate: 88 row: K column: 11
Seq primer: SP6
Class: BAC ends.
FEATURES
Location/Qualifiers
source
1..723
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-88K11"
/sex="Female"
/lab_host="DH10B"
/clone_lib="RPCI-23"
/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site 1:
EcoRI; Site 2: EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methylase. Size
selected DNA was cloned into the pBACe3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."
ORIGIN
Query Match          39.1%; Score 27; DB 8; Length 723;
Best Local Similarity 66.1%; Pred. No. 78;
Matches 39; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

Qy 8 ATACTGTAGGAGTTTCATTTTCACGAGGATATCAGACGGCTTTACTGGTAAATCACT 66
|||||
Db 444 ATACTGGAAGAGTCTCAAACTGACACGGCTGGCAAGATGGCTTAGCAGGTAAATGCT 386
|||||

RESULT 14
CL477612
LOCUS
DEFINITION
SAIL_276_F05.v1 SAIL Collection Arabidopsis thaliana genomic clone
SAIL_276_F05.v1, Genomic survey sequence.
ACCESSION
CL477612
VERSION
CL477612.1 GI:45944658

```

KEYWORDS GSS.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
REFERENCE 1 (bases 1 to 916)
AUTHORS Sessions, A., Burke, E., Presting, G., Aux, G., McElver, J., Patton, D.,
Dietrich, B., Ho, P., Bacwaden, J., Ko, C., Clarke, J. D., Cotton, D.,
Bullis, D., Snell, J., Miguel, T., Hutchison, D., Kimmerly, B.,
Mitzel, T., Katagiri, F., Glazebrook, J., Law, M. and Goff, S. A.
A high-throughput Arabidopsis reverse genetics system
Plant Cell 14 (12), 2985-2994 (2002)
22356987
MEDLINE
PUBMED 12468722
COMMENT Contact: Sessions A
Applied Trait Genetics
Syngenta Biotechnology Inc.
3054 Cornwallis Rd., Research Triangle Park, NC 27709, USA
Email: alien.sessions@syngenta.com
ABRC Stock Number CS812818; T-DNA left border flanking sequences of
Syngenta Arabidopsis Insertion Library (SAIL) lines are available
through the Arabidopsis Biological Resource Center (ABRC).
Sequences represent a pool of amplified genomic regions and not
single contiguous sequences.
Class: TDNA tagged.
Location/Qualifiers
1..916
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/ecotype="Columbia"
/db_xref="taxon:3702"
/clone="SAIL 276 F05.v1"
/clone_lib="SAIL Collection"
/note="T-DNA left border sequences were isolated using a
modified TAIL-PCR strategy"

FEATURES
source
Query Match 39.1%; Score 27; DB 9; Length 916;
Best Local Similarity 70.6%; Pred. No. 81;
Matches 36; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
Qy 19 GTTCAATTTCAGCAGGTATCAGAACGGCTTTACTGGTAATATACACTACA 69
|||||
Db 777 GTTGCCACTTCCCAAGGAACACACTGGCTTTAGCGGTGATACAGTACA 827
|||||

RESULT 15
BH998573/c
LOCUS oe123f09.b1 B.oleracea002 Brassica oleracea genomic, genomic survey
DEFINITION 701 bp DNA linear GSS 07-OCT-2002
sequence.
ACCESSION BH998573
VERSION BH998573.1 GI:23545088
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 701)
AUTHORS Delehaenty, K., Fewell, G., Fulton, L., McComb, W. R., Miner, T.,
Nash, W., Rabinowicz, P. D. and Wilson, R. K.
Whole genome shotgun reads from Brassica oleracea
Unpublished (2002)
Contact: Richard K. Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Plate: oe123 row: f column: 09
Seq primer: -21UpPOT forward
Class: shotgun
High quality sequence start: 39

FEATURES
source
High quality sequence stop: 551.
Location/Qualifiers
1..701
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/db_xref="taxon:3712"
/clone_lib="B.oleracea002"
/note="Vector: pOTw13; Whole genome shotgun library from
flowering buds. DNA was purified from a crude nuclear
prep using Brassica oleracea T01000DH3 buds provided by
Thomas Osborn at the University of Wisconsin.
DNA was provided by Pablo Rabinowicz (CSHL) and the
shotgun library prepared at Washington University Genome
Sequencing Center."

ORIGIN
Query Match 38.8%; Score 26.8; DB 8; Length 701;
Best Local Similarity 64.5%; Pred. No. 91;
Matches 40; Conservative 0; Mismatches 22; Indels 0; Gaps 0;
Qy 3 TGCTAATCTCTAGGAGTTTCAATTCAGCAGGGTATCAGAACGGCTTTACTGGTAATAT 62
|||||
Db 590 TTCTAATCTTTTATGAGTTTGATTATGAATATGATCAGAACGGCTTAATTGGTTTAA 531
|||||
Qy 63 CA 64
||
Db 530 CA 529

Search completed: November 20, 2005, 07:59:37
Job time : 1683 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 2005, 07:31:44 ; Search time 1443 Seconds
(without alignments)
2316.985 Million cell updates/sec

Title: US-10-780-250-4
Perfect score: 69
Sequence: 1 gatgctaatactgttagagt.....ttactgtaataatactaca 69

Scoring table: OLIGO NUC
Gapop '60.0', Gapext 60.0

Searched: 4708233 seqs, 24227607955 residues

Word size : 0

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	20	29.0	4490	3	AF481924	AF481924 Dictyoste
C 2	20	29.0	152799	2	AC150972	AC150972 Bos tauru
C 3	20	29.0	190527	2	AC119171	AC119171 Mus muscu
C 4	19	27.5	110000	2	AC128221_1	Continuation (2 of
C 5	19	27.5	157865	9	AC068413	AC068413 Homo sapi
C 6	19	27.5	165741	2	AC121063	AC121063 Rattus no
C 7	19	27.5	175406	10	AC131778	AC131778 Mus muscu
C 8	19	27.5	178528	10	AC131662	AC131662 Mus muscu
C 9	19	27.5	192595	2	AC127856	AC127856 Rattus no
C 10	19	27.5	219153	2	AC110939	AC110939 Rattus no
C 11	19	27.5	225253	2	AC098950	AC098950 Rattus no
C 12	19	27.5	277508	2	AC137325	AC137325 Rattus no
C 13	18	26.1	1659	6	AR551479	AR551479 Sequence
C 14	18	26.1	2162	10	BC083546	BC083546 Rattus no
C 15	18	26.1	2182	10	AF395746	AF395746 Rattus no
C 16	18	26.1	2470	10	AF387731	AF387731 Mus muscu
C 17	18	26.1	7568	6	AR091458	AR091458 Sequence
C 18	18	26.1	7568	6	AR104409	AR104409 Sequence
C 19	18	26.1	7568	6	AR260534	AR260534 Sequence

C 20	18	26.1	7568	6	AX743559	AX743559 Sequence
C 21	18	26.1	7568	14	SBVORFS	M89923 Sugarcane b
C 22	18	26.1	98246	2	AC146850_3	Continuation (4 of
C 23	18	26.1	110000	1	AE017225_14	Continuation (15 o
C 24	18	26.1	110000	1	AE017334_14	Continuation (15 o
C 25	18	26.1	110134	2	AC149660	AC149660 Bos tauru
C 26	18	26.1	154279	2	AC151077	AC151077 Bos tauru
C 27	18	26.1	160144	9	AC116340	AC116340 Homo sapi
C 28	18	26.1	172894	2	AC150571	AC150571 Bos tauru
C 29	18	26.1	179484	9	AC010595	AC010595 Homo sapi
C 30	18	26.1	185784	2	AC120860	AC120860 Mus muscu
C 31	18	26.1	187210	2	AC151057	AC151057 Bos tauru
C 32	18	26.1	188784	2	AC021318	AC021318 Homo sapi
C 33	18	26.1	201437	10	AL807236	AL807236 Mouse DNA
C 34	18	26.1	214170	2	AC118037	AC118037 Mus muscu
C 35	18	26.1	217924	2	AC105529	AC105529 Rattus no
C 36	18	26.1	222615	5	AL929345	AL929345 Zebrafish
C 37	18	26.1	223761	2	AC147195	AC147195 Bos tauru
C 38	18	26.1	230116	10	AL732620	AL732620 Mouse DNA
C 39	18	26.1	231559	2	AC096987	AC096987 Rattus no
C 40	18	26.1	244097	2	CR354394	CR354394 Danio rer
C 41	18	26.1	265423	2	AC128960	AC128960 Rattus no
C 42	18	26.1	265932	2	AC128425	AC128425 Rattus no
C 43	18	26.1	268324	5	AL954838	AL954838 Zebrafish
C 44	18	26.1	290117	1	AE017028	AE017028 Bacillus
C 45	18	26.1	301200	1	AP005374	AP005374 Thermosyn

ALIGNMENTS

RESULT 1	AF481924/c	4490 bp	DNA	linear	INV 19-AUG-2002
LOCUS	Dictyostelium discoideum cyclic GMP-binding protein D (gbpd) gene,				
DEFINITION	complete cds.				
ACCESSION	AF481924				
VERSION	AF481924.1	GI:21069540			
KEYWORDS	Dictyostelium discoideum				
SOURCE	Dictyostelium discoideum				
ORGANISM	Dictyostelium discoideum				
REFERENCE	1 (bases 1 to 4490)				
AUTHORS	Bosgraaf,L., Russcher,H., Smith,J.L., Wessels,D., Solis,D.R. and Van Haastert,P.J.M.				
TITLE	A novel cGMP signalling pathway mediating myosin phosphorylation and chemotaxis in Dictyostelium				
JOURNAL	EMBO J. (2002) In press				
REFERENCE	2 (bases 1 to 4490)				
AUTHORS	Goldberg,J.M., Bosgraaf,L., Van Haastert,P.J. and Smith,J.L.				
TITLE	Identification of four candidate cGMP targets in Dictyostelium				
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (10), 6749-6754 (2002)				
MEDLINE	22008080				
PUBMED	12011437				
REFERENCE	3 (bases 1 to 4490)				
AUTHORS	Smith,J.L., Goldberg,J.M., Bosgraaf,L., Russcher,H. and Van Haastert,P.J.M.				
TITLE	Direct Submission				
JOURNAL	Submitted (12-FEB-2002) Boston Biomedical Research Institute, 64 Grove St., Watertown, MA 02472, USA				
FEATURES	Location/Qualifiers				
source	1. 4490				
gene	/organism="Dictyostelium discoideum"				
mrna	/mol_type="genomic DNA"				
CDS	/db_xref="taxon:44689"				
	<57. .>4440				
	/gene="gbpd"				
	join(<57. .201,312. .471,575. .821,902. .1723,1876. .>4440)				
	/gene="gbpd"				
	/product="cyclic GMP-binding protein D"				
	join(57.201,312. .471,575. .821,902. .1723,1876. .4440)				
	/gene="gbpd"				
	/note="GbpD; RaeGEF"				


```

* 24228 24327: gap of unknown length
* 24328 28614: contig of 4287 bp in length
* 28615 28714: gap of unknown length
* 31677 31676: contig of 2962 bp in length
* 31677 44849: contig of 13073 bp in length
* 44850 44949: gap of unknown length
* 44950 53370: contig of 8421 bp in length
* 53371 53470: gap of unknown length
* 53471 55635: contig of 2165 bp in length
* 55636 55735: gap of unknown length
* 55736 74262: contig of 18527 bp in length
* 74263 74362: gap of unknown length
* 74363 77141: contig of 2779 bp in length
* 77142 77241: gap of unknown length
* 77242 80088: contig of 2847 bp in length
* 80089 80188: gap of unknown length
* 80189 87227: contig of 7039 bp in length
* 87228 87327: gap of unknown length
* 87328 89368: contig of 2041 bp in length
* 89369 89468: gap of unknown length
* 89469 93170: contig of 3702 bp in length
* 93171 93270: gap of unknown length
* 93271 101483: contig of 8213 bp in length
* 101484 101583: gap of unknown length
* 101584 103735: contig of 2152 bp in length
* 103736 103835: gap of unknown length
* 103836 108675: contig of 4840 bp in length
* 108676 108775: gap of unknown length
* 108776 115169: contig of 6394 bp in length
* 115170 115269: gap of unknown length
* 115270 117731: contig of 2462 bp in length
* 117732 117831: gap of unknown length
* 117832 121205: contig of 3374 bp in length
* 121206 121305: gap of unknown length
* 121306 129311: contig of 8006 bp in length
* 129312 129411: gap of unknown length
* 129412 133289: contig of 3878 bp in length
* 133290 133389: gap of unknown length
* 133390 141526: contig of 8137 bp in length
* 141527 141626: gap of unknown length
* 141627 144673: contig of 3047 bp in length
* 144674 144779: gap of unknown length
* 144779 152799: contig of 8026 bp in length.

```

FEATURES

```

source
1. 152799
/organism="Bos taurus"
/mol_type="genomic DNA"
/db_xref="taxon:9913"
/clone="CH240-35515"

```

ORIGIN

```

Query Match      29.0%; Score 20; DB 2; Length 152799;
Best Local Similarity 100.0%; Pred. No. 6.3;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```
Oy 7 AATACTGATGAGGATTTCAAT 26
|||||

```

```
Db 94315 AATACTGATGAGGATTTCAAT 94296
|||||

```

RESULT 3

```

AC119171
LOCUS AC119171 190527 bp DNA linear HTG 25-SEP-2004
DEFINITION Mus musculus chromosome 8 clone RP23-201B1 map 8, *** SEQUENCING IN
PROGRESS ***, 5 unordered pieces.
AC119171
AC119171.4 GI:51870831
VERSION HTG; HTGS PHASE1; HTGS FULLTOP; HTGS_ACTIVEFIN.
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

REFERENCE

```

AUTHORS
TITLE
JOURNAL
REFERENCE
Unpublished

```

REFERENCE

```

1 (bases 1 to 190527)
Birren,B., Nusbaum,C. and Lander,E.
Mus musculus chromosome 8, clone RP23-201B1
Unpublished
2 (bases 1 to 190527)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Babbitt,V., Bloom,T., Bougelavsky,L.,
Boukhalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Colangelo,M., Collins,K., Collymore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferrelira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gordon,S., Goyette,M., Graham,L., Grand-Pierre,N., Jones,C.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,K., Jones,C.,
Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,
MacLean,C., MacDonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollard,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J.J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (25-APR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 190527)
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V.,
Bloom,T., Bougelavsky,L., Boukhalter,B., Camarata,J., Chang,J.,
Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B.,
Dearellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L.,
Erickson,J., Faro,S., Ferrelira,P., FitzGerald,M., Gage,D.,
Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N.,
Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I.,
Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T.,
Lavine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,
MacLean,C., MacDonald,P., Major,J., Manning,J., Matthews,C.,
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C.,
Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R.,
Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (25-SEP-2004) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 3, 2004 this sequence version replaced gi:29135601.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@broad.mit.edu
----- Project Information
Center project name: L20405
Center clone name: 201_B_1
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.

```

* This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 21016: contig of 21016 bp in length
 * 21017 21116: gap of unknown length
 * 21117 40325: contig of 19209 bp in length
 * 40326 40425: gap of unknown length
 * 40426 50655: contig of 9640 bp in length
 * 50656 50165: gap of unknown length
 * 50166 127187: contig of 77022 bp in length
 * 127188 127287: gap of unknown length
 * 127288 190527: contig of 63240 bp in length.

FEATURES

source

1..190527
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="8"
 /map="8"
 /clone="RP23-201B1"
 /clone_lib="RPC1-23 Female Mouse BAC"

ORIGIN

Query Match 29.0%; Score 20; DB 2; Length 190527;
 Best Local Similarity 100.0%; Pred. No. 6;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 20 TTTCATTTCACGAGGGTAT 39

|||||
 Db 94411 TTTCATTTCACGAGGGTAT 94430

RESULT 4

AC128221_1/c

WPCOMMENT

Sequence split into 4 fragments LOCUS AC128221 Accession AC128221

Fragment Name	Begin	End
AC128221_0	1	110000
AC128221_1	100001	210000
AC128221_2	200001	310000
AC128221_3	300001	350171

Continuation (2 of 4) of AC128221 from base 100001 (AC128221 Rattus norvegicus clone CH2)

Query Match 27.5%; Score 19; DB 2; Length 110000;

Best Local Similarity 100.0%; Pred. No. 22;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 17 GAGTTTCAATTTCAGCAGG 35

|||||
 Db 96515 GAGTTTCAATTTCAGCAGG 96497

RESULT 5

AC068413/c

LOCUS AC068413 157865 bp DNA linear PRI 15-AUG-2002
 DEFINITION Homo sapiens chromosome 8, clone RP11-27319, complete sequence.

ACCESSION AC068413

VERSION AC068413.11 GI:22218575

KEYWORDS HTG.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 157865)

Birren,B., Nussbaum,C. and Lander,E.

Homo sapiens chromosome 8, clone RP11-27319

Unpublished

2 (bases 1 to 157865)

Birren,B., Linton,L., Nussbaum,C., Lander,E., Abraham,H., Allen,N.,

Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,

Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,

Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,

Collamore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,

Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
 Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
 Grand-Pierre,N., Grant,G., Hagos,B., Hearford,A., Horton,L.,
 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
 Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J.,
 Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
 McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPeeters,R.,
 Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
 Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
 O'Neill,D., Olivari,T.M., Oliver,J., Peterson,K., Pierre,N.,
 Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
 Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
 Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
 Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
 Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Direct Submission

Submitted (02-MAY-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE

AUTHORS

3 (bases 1 to 157865)

Birren,B., Nussbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
 Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,
 Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collamore,A.,
 Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
 Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
 Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
 Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
 Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
 Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
 McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
 Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
 O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
 Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
 Roman,J., Roy,A., Schauer,S., Schuback,R., Seaman,S., Severy,P.,
 Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
 Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
 Zembek,L., Zimmer,A. and Zody,M.

Direct Submission

Submitted (14-AUG-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA

4 (bases 1 to 157865)

Birren,B., Nussbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
 Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,
 Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collamore,A.,
 Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
 Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
 Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
 Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
 Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
 Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
 McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
 Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
 O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
 Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
 Roman,J., Roy,A., Schauer,S., Schuback,R., Seaman,S., Severy,P.,
 Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
 Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
 Zembek,L., Zimmer,A. and Zody,M.

Direct Submission

Submitted (15-AUG-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA

On Aug 14, 2002 this sequence version replaced gi:18643472.

All repeats were identified using RepeatMasker:

Smith, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WBIR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L10308
Center clone name: 273_I_9

FEATURES

Source

Location/Qualifiers

```

1. .157865
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="8"
/map="8"
/clone="RP11-27319"
/clone.lib="RPC1-11 Human Male BAC"
280. .404
/rpt_family="CT-rich"
406. .583
/rpt_family="TTTC"n"
2103. .2147
/rpt_family="CA)n"
complement(2459. .2847)
/rpt_family="LIME"
3415. .3579
/rpt_family="MIR3"
complement(4240. .5444)
/rpt_family="L1PA3"
5436. .6665
/rpt_family="L1PA3"
6740. .6892
/rpt_family="MIR"
7370. .7535
/rpt_family="MIR3"
7536. .7568
/rpt_family="TCTCC)n"
7755. .7790
/rpt_family="MLT1J2"
7791. .7854
/rpt_family="TTATA)n"
7855. .8087
/rpt_family="MLT1J2"
8246. .8276
/rpt_family="AT-rich"
8334. .8746
/rpt_family="MSTA"
9534. .9632
/rpt_family="MER20"
9861. .10397
/rpt_family="MLT2F"
10937. .10959
/rpt_family="AT-rich"
complement(11349. .12239)
/rpt_family="L1M4"
complement(12240. .13722)
/rpt_family="L1M4"
complement(13723. .13917)
/rpt_family="MLT1F"
complement(13918. .14195)
/rpt_family="MLT1F-int"
14281. .14709
/rpt_family="L1PB1"
14714. .19896
/rpt_family="L1PB1"
19897. .19967
/rpt_family="TA)n"
19972. .20323
/rpt_family="L1PB1"
complement(20352. .20502)
/rpt_family="MLT1F-int"
complement(20601. .20789)
/rpt_family="MLT1F-int"
21641. .22011
/rpt_family="MLT1A1"
complement(22228. .22344)
/rpt_family="L2"
complement(22805. .22855)

```

```

repeat_region /rpt_family="L3"
22880. .22934
/rpt_family="L2"
25615. .26058
/rpt_family="MLT1C"
26688. .26771
/rpt_family="L2"
27056. .28104
/rpt_family="L1PA15"
28296. .28594
/rpt_family="AluSx"
28595. .28633
/rpt_family="TA)n"
complement(29354. .29515)
/rpt_family="MIR"
29587. .29694
/rpt_family="CT-rich"
complement(29695. .29747)
/rpt_family="L2"
complement(30480. .30638)
/rpt_family="FAM"
31028. .31395
/rpt_family="THE1C"
31396. .31674
/rpt_family="THE1C-int"
31675. .31699
/rpt_family="(T)n"
complement(31704. .31896)
/rpt_family="MIR"
32818. .32840
/rpt_family="AT-rich"
complement(32936. .33022)
/rpt_family="MIR"
complement(33723. .33802)
/rpt_family="Charlie8"

```

Query Match 27.5%; Score 19; DB 9; Length 157865;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 TAATACGTAGGAGTTTCA 24

Db 85903 TAATACGTAGGAGTTTCA 85885

RESULT 6

AC121063/c

AC121063 Rattus norvegicus clone CH230-337G6, *** SEQUENCING IN PROGRESS
DEFINITION ***

ACCESSION AC121063

VERSION AC121063.4 GI:23907858

KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ENRICHED.

SOURCE Rattus norvegicus (Norway rat)

ORGANISM

Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 165741)

AUTHORS

Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Ayagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Blawie,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Cesar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,

Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregorgis, E., Geier, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, C., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpachy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovat, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuhewa, L., Loulseghe, H., Lozado, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokemele, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, I., Pfannkuch, C., Popper, F., Poidexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Vallas, R., Vera, V., Villanana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstein, G., and Gibbs, R.A.

Direct Submission
Unpublished
2 (bases 1 to 165741)
Worley, K.C.
Direct Submission
Submitted (15-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 165741)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (12-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Oct 12, 2002 this sequence version replaced gi:21313694.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GYDS
Center clone name: CH230-337G6
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 154673 bases at least Q40

Consensus quality: 155599 bases at least Q30
Consensus quality: 156415 bases at least Q20
Estimated insert size: 160958; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently consists of 1 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 165741: contig of 165741 bp in length.

FEATURES
source
1..165741
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-337G6"
misc_feature
1..2052
/note="wgs end extension
clone_end:Sp6"
misc_feature
2103..3192
/note="wgs end extension
clone_end:Sp6"
misc_feature
complement(4262..5151)
/note="clone boundary
clone_end:Sp6
site:Mbol
end_sequence:RXAQB39TV"
misc_feature
161745..163531
/note="wgs contig"
misc_feature
164580..165416
/note="clone boundary
clone_end:T7
site:Mbol
end_sequence:RXAQB39TJ"

ORIGIN
Query Match 27.5%; Score 19; DB 2; Length 165741;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 10 ACTGTAGGAGTTTCAATT 28
|||||
Db 47960 ACTGTAGGAGTTTCAATT 47942
|||||
RESULT 7
AC131778/c
LOCUS AC131778 175406 bp DNA linear ROD 15-MAY-2004
DEFINITION Mus musculus BAC clone RP24-328C2 from chromosome 7, complete
AC131778
AC131778
VERSION AC131778.3 GI:46275987
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 175406)
AUTHORS Wang, C., Bielicki, L. and Meyer, R.
TITLE The sequence of Mus musculus BAC clone RP24-328C2
JOURNAL Unpublished (2001)
REFERENCE
2 (bases 1 to 175406)
AUTHORS McPherson, J.D. and Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (25-AUG-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT


```

/rpt_family="MIR"
48059. .48257
/rpt_family="L1"
48726. .48920
/rpt_family="B4"
48958. .49004
/rpt_family="B4"
50214. .50836
/rpt_family="L1"
54064. .54132
/rpt_family="Alu"
56139. .56247
/rpt_family="ERVK"
56858. .57079
/rpt_family="B2"
57979. .58056
/rpt_family="L1"
58057. .58161
/rpt_family="Alu"
59291. .59382
/rpt_family="ERVK"
59411. .59690
/rpt_family="B4"
60654. .61915
/rpt_family="L1"
61957. .62217
/rpt_family="B4"
65286. .65414
/rpt_family="Alu"
65406. .65530
/note="Unresolved simple sequence repeat."
66468. .66747
/rpt_family="B4"
67287. .67337
/rpt_family="MER1_type"

Query Match      27.5%; Score 19; DB 10; Length 175406;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      26  TTTTCAGCAGGGTATCAGAA 44
Db      33431 TTTTCAGCAGGGTATCAGAA 33413

RESULT 8
AC131662/c
LOCUS      AC131662      178528 bp      DNA      linear      ROD 01-FEB-2004
DEFINITION Mus musculus chromosome 7 clone RP23-257H24, complete sequence.
ACCESSION  AC131662
VERSION     AC131662.4  GI:41529855
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS     Wilsoan,R.K.
TITLE       The sequence of Mus musculus clone
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 178528)
AUTHORS     McPherson,J.D. and Waterston,R.H.
TITLE       Direct Submission
JOURNAL     Submitted (25-AUG-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 178528)
AUTHORS     Wilsoan,R.K.
TITLE       Direct Submission
JOURNAL     Submitted (08-NOV-2003) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   4  (bases 1 to 178528)
AUTHORS     Wilsoan,R.K.
TITLE       Direct Submission

Submitted (01-FEB-2004) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
On Feb 1, 2004 this sequence version replaced gi:38229414.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site:http://genome.wustl.edu
Contact: submissions@watson.wustl.edu
----- Project Information -----
Center project name: M_BA0257H24
-----
FEATURES             Location/Qualifiers
     source            1..178528
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="7"
                     /clone="RP23-257H24"
ORIGIN
Query Match      27.5%; Score 19; DB 10; Length 178528;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      26  TTTTCAGCAGGGTATCAGAA 44
Db      153670 TTTTCAGCAGGGTATCAGAA 153652

RESULT 9
AC127856/c
LOCUS      AC127856      192595 bp      DNA      linear      HTG 12-OCT-2002
DEFINITION Rattus norvegicus clone CH230-255D13, WORKING DRAFT SEQUENCE.
ACCESSION  AC127856
VERSION     AC127856.2  GI:23907817
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
ORGANISM    Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.

REFERENCE
AUTHORS     1  (bases 1 to 192595)
            Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Centek,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denison,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisl,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,S., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,

```

Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwackeme, O., Okwono, G., Olarpunsaogon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankoch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.-J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleciyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

1 192595: contig of 192595 bp in length.

Location/Qualifiers

1. 192595

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/db_xref="taxon:10116"

/clone="CH230-255D13"

1. 1502

/note="wgs end extension"

clone_end:T7"

8610..9400

/note="clone boundary"

clone_end:T7"

site:MboI

end sequence:RXACK19TVB"

73968..108480

/note="clone boundary"

clone_end:Sp6"

191249..192595

/note="wgs end extension"

clone_end:Sp6"

end sequence:RXACK19TVB"

184275..186004

/note="wgs end extension"

clone_end:Sp6"

191249..192595

/note="wgs end extension"

clone_end:Sp6"

end sequence:RXACK19TVB"

184275..186004

/note="wgs end extension"

clone_end:Sp6"

191249..192595

/note="wgs end extension"

clone_end:Sp6"

end sequence:RXACK19TVB"

184275..186004

/note="wgs end extension"

clone_end:Sp6"

191249..192595

/note="wgs end extension"

clone_end:Sp6"

end sequence:RXACK19TVB"

184275..186004

/note="wgs end extension"

clone_end:Sp6"

191249..192595

/note="wgs end extension"

clone_end:Sp6"

end sequence:RXACK19TVB"

184275..186004

/note="wgs end extension"

clone_end:Sp6"

191249..192595

/note="wgs end extension"

clone_end:Sp6"

end sequence:RXACK19TVB"

184275..186004

/note="wgs end extension"

clone_end:Sp6"

191249..192595

/note="wgs end extension"

clone_end:Sp6"

end sequence:RXACK19TVB"

184275..186004

/note="wgs end extension"

clone_end:Sp6"

191249..192595

/note="wgs end extension"

clone_end:Sp6"

end sequence:RXACK19TVB"

184275..186004

/note="wgs end extension"

clone_end:Sp6"

191249..192595

/note="wgs end extension"

clone_end:Sp6"

end sequence:RXACK19TVB"

184275..186004

/note="wgs end extension"

clone_end:Sp6"

191249..192595

FEATURES

source

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

1 192595: contig of 192595 bp in length.

Location/Qualifiers

1. 192595

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/db_xref="taxon:10116"

/clone="CH230-255D13"

1

Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuewa, L., Loulseged, H., Lozado, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwakoelameh, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, P., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, K., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlecyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Direct Submission
Unpublished
2 (bases 1 to 219353)
Worley, K.C.

Direct Submission
Submitted (17-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 219353)
Rat Genome Sequencing Consortium.

Direct Submission
Submitted (09-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Sep 19, 2002 this sequence version replaced gi:22450400.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). As a result, the sequence may extend beyond the ends of the clone and there may be contigs that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GQGG
Center clone name: CH230-238K5
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 190762 bases at least Q40
Consensus quality: 193329 bases at least Q30
Consensus quality: 195199 bases at least Q20
Estimated insert size: 216488; sum-of-contigs estimation
Quality coverage: 3x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently consists of 2 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence

* as soon as it is available and the accession number will
* be preserved
* 1 208973: contig of 208973 bp in length
* 208974 209073: gap of unknown length
* 209074 219353: contig of 10280 bp in length.

FEATURES
Location/Qualifiers
source
1..219353
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clones="CH230-238K5"
1..1879
/note="wgs contig"
misc_feature
4520..6523
/note="wgs contig"
misc_feature
209074..210405
/note="wgs contig"
ORIGIN
Query Match 27.5% Score 19; DB 2; Length 219353;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 10 ACTGTAGGAGTTTCAATT 28
|||||
Db 165317 ACTGTAGGAGTTTCAATT 165299
|||||

RESULT 11
AC098950/c
LOCUS
DEFINITION Rattus norvegicus clone CH230-175D13, *** SEQUENCING IN PROGRESS
***, 2 unordered pieces.
AC098950
VERSION AC098950.5 Gi:30581567
HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE
1 (bases 1 to 225253)
Muzny, D., Marté, Metzker, M., Lee, A., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alabrooks, S., Amin, A., Anguiano, D., Anyalbechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuewa, L., Loulseged, H., Lozado, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,

Nwaokemele, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkoch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D., Smeile, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villanana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleciyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Direct Submission
Unpublished
2 (bases 1 to 225253)
Worley, K.C.

Direct Submission
Submitted (07-NOV-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 225253)
Rat Genome Sequencing Consortium.

Direct Submission
Submitted (13-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On May 13, 2003 this sequence version replaced gi:22855283.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GJGF
Center clone name: CH230-175D13
----- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 215069 bases at least Q40
Consensus quality: 217196 bases at least Q30
Consensus quality: 218747 bases at least Q20
Estimated insert size: 230068; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently consists of 2 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

* 1 223942: contig of 223942 bp in length
* 223943 224042: gap of unknown length

* 224043 225253: contig of 1211 bp in length.

FEATURES
source
Location/Qualifiers
1. .225253
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clones="CH230-175D13"
1. .2133
/note="wgs contig"
misc_feature
221378..222396
/note="wgs contig"
misc_feature
222447..223942
/note="wgs contig"

ORIGIN
Query Match 27.5% Score 19; DB 2; Length 225253;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 17 GAGTTTCAATTTCAGCAGG 35
Db 126374 GAGTTTCAATTTCAGCAGG 126356
|||||
AC137325 277508 bp DNA linear HTG 20-NOV-2002
Rattus norvegicus clone CH230-unknown, *** SEQUENCING IN PROGRESS
*** 7 unordered pieces.
AC137325
AC137325.1 GI:25138410
HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 277508)
Muzny, D., Marle, Metzker, M., Lee, Abramzon, S., Adams, C., Alder, J., Allen, C., Allen, H., Alebrooks, S., Amin, A., Anguiano, D., Ayalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Cesar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M.B., Davis, C., Davy-Carroil, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hollins, B., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Jackson, L., Jackson, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuhewa, L., Loulseghe, H., Lozada, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokemele, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkoch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,

Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Shen, H., Sanders, M., Savary, G., Scherer, S., Scott, G., Shatman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, K., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, X., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Direct Submission
Unpublished
2 (bases 1 to 277508)
Rat Genome Sequencing Consortium.
Submitted (20-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: K2RY

Center clone name: CH230-unknown

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 204136 bases at least Q40

Consensus quality: 209148 bases at least Q30

Consensus quality: 212449 bases at least Q20

Estimated insert size: 212734; sum-of-contigs estimation

Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently consists of 7 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

* 1
* 63395: contig of 63395 bp in length
* 63396
* 63495: gap of unknown length
* 63496
* 269704: contig of 206209 bp in length
* 269705
* 269804: gap of unknown length
* 269805
* 270862: contig of 1058 bp in length
* 270863
* 270962: gap of unknown length
* 270963
* 272445: contig of 1483 bp in length
* 272446
* 272445: gap of unknown length
* 272546
* 274059: contig of 1514 bp in length
* 274060
* 274159: gap of unknown length
* 274160
* 275289: contig of 1130 bp in length
* 275290
* 275389: gap of unknown length
* 275390
* 277508: contig of 2119 bp in length.

FEATURES

source
1..277508
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-unknown"
19553..21331
/note="wgs_contig"

misc_feature

77484..78741
/note="wgs_contig"

misc_feature

146795..148635
/note="wgs_contig"

ORIGIN

Query Match 27.5%; Score 19; DB 2; Length 277508;

Best Local Similarity 100.0%; Pred. No. 18;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 ACTGTAGGAGTTTCAATT 28
|||||

Db 90537 ACTGTAGGAGTTTCAATT 90519
|||||

RESULT 13

AR551479

LOCUS AR551479 1659 bp DNA linear PAT 08-OCT-2004

DEFINITION Sequence 6610 from patent US 6747137.

ACCESSION AR551479

VERSION AR551479.1 GI:53944654

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 1659)

AUTHORS Weinstock, K.G. and Bush, D.

TITLE Nucleic acid sequences relating to Candida albicans for diagnostics and therapeutics

JOURNAL Patent: US 6747137-A 6610 08-JUN-2004;

FEATURES Location/Qualifiers

source

1..1659

/organism="unknown"

/mol_type="genomic DNA"

ORIGIN

Query Match 26.1%; Score 18; DB 6; Length 1659;

Best Local Similarity 100.0%; Pred. No. 1.5e+02;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 49 TTTACTGTGTAATATCACT 66
|||||

Db 682 TTTACTGTGTAATATCACT 699
|||||

RESULT 14

BC083546/c

LOCUS BC083546 2162 bp mRNA linear ROD 12-OCT-2004

DEFINITION Rattus norvegicus SNM1-like, mRNA (cDNA clone MGC:93142 IMAGE:7134940), complete cds.

ACCESSION BC083546

VERSION BC083546.1 GI:54035285

KEYWORDS MGC.

SOURCE Rattus norvegicus (Norway rat)

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

REFERENCE 1 (bases 1 to 2162)

AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,

Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S.A., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kerteman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skaleka, U., Smailus, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.

Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

12477932

2 (bases 1 to 2162)

Director MGC Project.

Direct Submission

Submitted (01-OCT-2004) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgapbs-remail.nih.gov

Tissue Procurement: Howard Jacobs

cDNA Library Preparation: Express Genomics

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305

Web site: <http://www-shgc.stanford.edu>

Contact: (Dickson, Mark) mcd@paxil.stanford.edu

Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAK Plate: 185 Row: m Column: 18

This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 22129750.

FEATURES

source

Location/Qualifiers

1..2162

/organism="Rattus norvegicus"

/mol_type="mRNA"

/db_xref="taxon:10116"

/clone="MGC:93142 IMAGE:7134940"

/tissue_type="Testis, rat (Brown Norway)"

/clone_lib="NIH MGC_238"

/lab_host="DH10B"

/note="Vector: pExpress1"

gene

1..2162

/gene="Snm1l"

/db_xref="LocusID:259171"

/db_xref="RGD:708574"

CDS

24..2120

/gene="Snm1l"

/codon_start=1

/product="SNM1-like"

/protein_id="AAH83546.1"

/db_xref="GI:54035286"

/db_xref="LocusID:259171"

/db_xref="RGD:708574"

/translation="MSSFGQOMEYPTISIDFRENKARAYFLSHCHDKHMKGLRA
PSMKRLECSLKVFYCSVTKELLITSPKTKFWENRIIAIEITPTQVSLVDSASGE
KEEVVTLTPAGHCPGSMVFLQGSNGTVLTGDFRLAKGEVSRMELHSGGRVKDIO
SVYLDITFCDFRFYQIPSEBECLRGVLEVRSWITRSPKHVWMLNCKAAYGEYLFN
LSEELGVQVHVDKIDMPDILHLITDRNTQIHACRHPKAEYFQWNKLPQGMAS
KTKVLTHTISIKPSTWFGTRKTNVITRGESSYRACFSFHSYSSEIKDFLSYICP
VNAYPNVIPGLTVDKWDFLPLCRSQCAEPKYKPLGKLKARTVHLDSSEDDLLF
DDPLTHSRKRVQVTLHPVEFSMKALPLDQPELGSPGCCAKESFPSLANFVDC
DESNDSGELETPSLQGLGPTLLPQGNADPDVPRWEVFFRKRDEITDECLENL
PSSIETGSGSPKRFSDSPKLGSDSGESTHISQNSQSTHITDQSGQWDSQCDTV

LLSQEKSGDSTSLNKDITYKPKDQSIASQIQNALCPQDTHCDLKSQAEVNGVPC
IEPDTVSGRKSPEKTSLSLTQADSSQSSDFEIPSTPEALPKPHEHLQFLYGLKATG
ESIVLKKENVHSQIFK"

ORIGIN

Query Match 26.1%; Score 18; DB 10; Length 2162;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 GTAGGAGTTTCAATTTC 30
|||||
Db 295 GTAGGAGTTTCAATTTC 278
|||||

RESULT 15

AF395746/c 2182 bp mRNA linear ROD 31-JUL-2002

LOCUS AF395746 Rattus norvegicus SNM1-like protein mRNA, complete cds.

DEFINITION AF395746

ACCESSION AF395746

VERSION AF395746.1 GI:22023556

KEYWORDS

SOURCE Rattus norvegicus (Norway rat)

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE

1 (bases 1 to 2182)
Li, L., Zhou, Y., Xie, G. and Cowan, M.J.
The mouse and rat SNM1-like genes, cloning, expression and mapping
JOURNAL Unpublished

2 (bases 1 to 2182)
Li, L., Zhou, Y. and Cowan, M.J.
Direct Submission
JOURNAL Submitted (27-JUN-2001) Pediatrics, University of California San Francisco, 513 Parnassus Avenue, San Francisco, CA 94143, USA

FEATURES

source

1..2182

/organism="Rattus norvegicus"

/mol_type="mRNA"

/db_xref="taxon:10116"

56..2149

/note="SNM1l"

/codon_start=1

/product="SNM1-like protein"

/protein_id="AAM89124.1"

/db_xref="GI:22023557"

/translation="MSSFGQOMEYPTISIDFRENKARAYFLSHCHDKHMKGLRA
PSMKRLECSLKVFYCSVTKELLITSPKTKFWENRIIAIEITPTQVSLVDSASGE
KEEVVTLTPAGHCPGSMVFLQGSNGTVLTGDFRLAKGEVSRMELHSGGRVKDIO
SVYLDITFCDFRFYQIPSEBECLRGVLEVRSWITRSPKHVWMLNCKAAYGEYLFN
LSEELGVQVHVDKIDMPDILHLITDRNTQIHACRHPKAEYFQWNKLPQGMAS
KTKVLTHTISIKPSTWFGTRKTNVITRGESSYRACFSFHSYSSEIKDFLSYICP
VNAYPNVIPGLTVDKWDFLPLCRSQCAEPKYKPLGKLKARTVHLDSSEDDLLF
DDPLTHSRKRVQVTLHPVEFSMKALPLDQPELGSPGCCAKESFPSLANFVDC
DESNDSGELETPSLQGLGPTLLPQGNADPDVPRWEVFFRKRDEITDECLENL
PSSIETGSGSPKRFSDSPKLGSDSGESTHISQNSQSTHITDQSGQWDSQCDTV
LLSQEKSGDSTSLNKDITYKPKDQSIASQIQNALCPQDTHCDLKSQAEVNGVSC
IEPDTVSGRKSPEKTSLSLTQADSSQSSDFEIPSTPEALPKPHEHLQFLYGLKATG
ESIXVEKRKCSLPDI"

ORIGIN

Query Match 26.1%; Score 18; DB 10; Length 2182;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 GTAGGAGTTTCAATTTC 30
|||||

Db 327 GTAGGAGTTTCAATTTC 310
|||||

Search completed: November 20, 2005, 08:44:22
Job time : 1462 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 2005, 07:59:44 ; Search time 1786 Seconds
(without alignments)
1470.568 Million cell updates/sec

Title: US-10-780-250-4

Perfect score: 69
Sequence: 1 gatgctaatactgttagagt.....ttactgtaatactactaca 69

Scoring table: OLIGO NUC

Gapop 60.0 , Gapext 60.0

Searched: 34239544 seqs, 19032134700 residues

Word size : 0

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:*

1: gb_est1:*

2: gb_est2:*

3: gb_hc:*

4: gb_est3:*

5: gb_est4:*

6: gb_est5:*

7: gb_est6:*

8: gb_gse1:*

9: gb_gse2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	29.0	541	8	BH058722 RPCI-24-9
C 2	20	29.0	652	4	BJ395667 BJ395667
C 3	20	29.0	718	7	CN233012 WLB055A10
C 4	19	27.5	855	9	CC905007 t018o13ba
C 5	18	26.1	202	4	B1424553 sah52902.
6	18	26.1	221	9	CS590287 tigr-gss-
7	18	26.1	297	2	AW311778 5665 MARC
8	18	26.1	315	2	AW311777 5664 MARC
C 9	18	26.1	340	2	BB847064 BB847064
C 10	18	26.1	340	5	BY306203 BY306203
C 11	18	26.1	361	6	BY774046 BY774046
C 12	18	26.1	385	5	BY166908 BY166908
C 13	18	26.1	392	5	BY095388 BY095388
C 14	18	26.1	395	5	BY158605 BY158605
C 15	18	26.1	408	1	AA140626 MEEG0147.
16	18	26.1	419	2	AW311776 5663 MARC
C 17	18	26.1	429	4	B1468631 sai01dl1.
C 18	18	26.1	462	9	CC765669 CH240_130
C 19	18	26.1	484	7	CR450990 CR450990
C 20	18	26.1	502	2	BB474254 sp61b04.Y
C 21	18	26.1	526	6	CB720000 AMGNNUC:C
C 22	18	26.1	557	4	BM093229 saj06e07.
C 23	18	26.1	565	4	B1973331 sai186h11.
C 24	18	26.1	568	7	CO618344 DG9-185a5

C 25	18	26.1	585	7	CN697471	CN697471	E0394G01-
C 26	18	26.1	602	2	BB630242	BB630242	BB630242
C 27	18	26.1	613	2	BB660218	BB660218	BB660218
C 28	18	26.1	618	7	CF083776	CF083776	QHL20J03.
C 29	18	26.1	618	8	AZ973007	AZ973007	2M0247001
C 30	18	26.1	628	2	BB659504	BB659504	BB659504
C 31	18	26.1	634	7	CK768923	CK768923	Gm-r1030-
C 32	18	26.1	649	2	BB635523	BB635523	BB635523
C 33	18	26.1	652	2	BB634316	BB634316	BB634316
C 34	18	26.1	654	6	BY729142	BY729142	BY729142
C 35	18	26.1	657	2	BB589727	BB589727	BB589727
C 36	18	26.1	659	6	BY734927	BY734927	BY734927
C 37	18	26.1	668	6	BY747058	BY747058	BY747058
C 38	18	26.1	674	2	BB589750	BB589750	BB589750
C 39	18	26.1	698	7	CO681677	CO681677	DG11-132a
C 40	18	26.1	725	2	BB632064	BB632064	BB632064
C 41	18	26.1	739	9	CC570244	CC570244	CH240_445
C 42	18	26.1	762	9	CC924394	CC924394	t074i24ba
C 43	18	26.1	822	7	CK597490	CK597490	CK597490
C 44	18	26.1	871	9	CC510566	CC510566	CH240_353
C 45	18	26.1	884	7	CO402870	CO402870	AGENCOURT

ALIGNMENTS

RESULT 1
BH058722 541 bp DNA linear GSS 18-JUL-2001
LOCUS RPCI-24-9406.TJB RPCI-24 Mus musculus genomic clone RPCI-24-9406,
DEFINITION Genomic survey sequence.
ACCESSION BH058722
VERSION BH058722.1 GI:14868029
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 541)
AUTHORS Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregorgis,E.,
Russell,B., de Jong,P. and Fraser,C.M.
TITLE Mouse BAC End Sequences from Library RPCI-24
JOURNAL Unpublished (1999)
COMMENT Other GSSs: RPCI-24-9406.TJ RPCI-24-9406.TV RPCI-24-9406.TVB
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC
library availability, please contact Pieter de Jong
(pdejong@mail.cho.org). Clones may be purchased from BACPAC
Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end
page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 94 row: 0 Column: 6
Seq primer: SP6
Class: BAC ends.

Location/Qualifiers

1. 541
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-9406"
/sex="Male"
/cell_type="Spleen/Brain"
/clone_lib="RPCI-24"
/note="Vector: pTARBAC1, Site 1: BamH1; Site 2: BamH1;
RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the pTARBAC1 cloning vector at the

FEATURES

source
1. 541
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-9406"
/sex="Male"
/cell_type="Spleen/Brain"
/clone_lib="RPCI-24"
/note="Vector: pTARBAC1, Site 1: BamH1; Site 2: BamH1;
RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the pTARBAC1 cloning vector at the

BamHI sites using MboI partially digested male C57BL/6J DNA."

ORIGIN

Query Match 29.0%; Score 20; DB 8; Length 541;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 20 TTTCATTTTCAGCAGGGTAT 39
|||||
Db 433 TTTCATTTTCAGCAGGGTAT 452

RESULT 2

BJ395667/c
LOCUS BJ395667 Dictyostelium discoideum cDNA library, SF Dictyostelium
DEFINITION dictoideum cDNA clone dds39p17 5', mRNA sequence.

ACCESSION BJ395667

VERSION BJ395667.1 GI:19306753

KEYWORDS EST.

SOURCE Dictyostelium discoideum

ORGANISM Dictyostelium discoideum

REFERENCE 1 (bases 1 to 652)

AUTHORS Urushihara, H., Tanaka, Y., Kohara, Y. and Shin-i, T.

TITLE Full length cDNA of Dictyostelium discoideum at the slug stage

JOURNAL Unpublished (2002)

COMMENT Contact: Tadao Shin-i

Center For Genetic Resource Information

National Institute of Genetics

1111 Yata, Mishima, Shizuoka 411-8540, Japan

Tel: 81-559-81-6856

Fax: 81-559-81-6855

Email: tshini@genes.nig.ac.jp.

Location/Qualifiers

FEATURES

source
1..652
/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="dds39p17"
/sex="mat A"
/dev_stage="Slug stage"
/clone_lib="Dictyostelium discoideum cDNA library, SF"

ORIGIN

Query Match 29.0%; Score 20; DB 4; Length 652;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 48 CTTTACTCGTAATATCACTA 67
|||||
Db 147 CTTTACTCGTAATATCACTA 128

RESULT 3

CN233012/c
LOCUS CN233012 718 bp mRNA linear EST 09-APR-2004
DEFINITION WL8055A10.ab1 Wttestis Gallus gallus cDNA 5', mRNA sequence.

ACCESSION CN233012

VERSION CN233012.1 GI:46336756

KEYWORDS EST.

SOURCE Gallus gallus (chicken)

ORGANISM Gallus gallus

REFERENCE 1 (bases 1 to 718)

AUTHORS Archosauria; Aves; Neognathae; Galliformes; Phasianidae;

Phasianinae; Gallus.

1 (bases 1 to 718)

Savolainen, P., Fitzsimmons, C.J., Arvestad, L., Andersson, L. and

Lundberg, J.

TITLE EST analysis of brain and testis cDNA libraries from White Leghorn

and Red Jungle Fowl

JOURNAL COMMENT

Unpublished (2004)
Contact: Peter Savolainen
Department of Biotechnology
Royal Institute of Technology, KTH
SE-106 91 Stockholm, SWEDEN
Tel: +46 (0)8 5537 8481
Fax: +46 (0)8 5537 8335
Email: Peter.Savolainen@biotech.kth.se
Seq primer: M13 reverse primer.

FEATURES

source
1..718
Location/Qualifiers
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="White Leghorn"
/db_xref="taxon:9031"
/sex="male"
/lab_host="ElectroMAX DH10B (Invitrogen)"
/clone_lib="Wttestis"
/note="Organ: testis; Vector: pSPORT-1; Site 1: Hind III;
Site 2: EcoRI; The cDNA libraries were created with the
SuperScript Plasmid System (Invitrogen)."

ORIGIN

Query Match 29.0%; Score 20; DB 7; Length 718;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 15 AGGAGTTTCAATTTTCAGCAG 34

|||||

Db 600 AGGAGTTTCAATTTTCAGCAG 581

RESULT 4

CC905007/c

LOCUS CC905007 855 bp DNA linear GSS 08-AUG-2003

DEFINITION t0180i3ba.r1 TAMBT Bos taurus genomic clone t0180i3ba, genomic survey sequence.

ACCESSION CC905007

VERSION CC905007.1 GI:33523940

KEYWORDS GSS.

SOURCE Bos taurus (cow)

ORGANISM Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovinae; Bos.

1 (bases 1 to 855)

Lin, S., Najjar, F.Z., Adelson, D., Gill, C.A. and Roe, B.A.

Bovine BAC End Sequences from Library TAMBT

Unpublished (2003)

JOURNAL

COMMENT

Contact: Bruce A. Roe

Advanced Center for Genome Technology

University of Oklahoma Department of Chemistry and Biochemistry

620 Parrington Oval, Room 208, Norman, OK 73019, USA

Tel: 405 325 4912

Fax: 405 325 7762

Email: broe@ou.edu

Class: BAC ends

High quality sequence start: 78

High quality sequence stop: 501.

FEATURES

source
1..855
Location/Qualifiers
/organism="Bos taurus"
/mol_type="genomic DNA"
/strain="Angus bull T A M U Shoshone Y6 11519666"
/db_xref="taxon:9913"
/clone="t0180i3ba"
/sex="Male"
/cell_type="Blood"
/clone_lib="TAMBT"
/note="Vector: pBeloBAC11; Site 1: HindIII; Site 2:
HindIII; TAMBT Bovine BAC library (Male) produced by Texas
A&M University, Department of Animal Science."

ORIGIN

Query Match 27.5%; Score 19; DB 9; Length 855;
 Best Local Similarity 100.0%; Pred. No. 6.1;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 16 GGAGTTCAATTTCAGCAG 34
 |||||
 Db 339 GGAGTTCAATTTCAGCAG 321

RESULT 5
 BI424553/C
 LOCUS BI424553 202 bp mRNA linear EST 21-JUL-2004
 DEFINITION sahs52g02.y1 Gm-cl036 Glycine max cDNA clone GENOME SYSTEMS CLONE
 ID: Gm-cl036-5283 5', mRNA sequence.
 ACCESSION BI424553
 VERSION BI424553.1 GI:15200829
 KEYWORDS EST.
 SOURCE Glycine max (soybean)
 ORGANISM Glycine max

REFERENCE 1 (bases 1 to 202)
 AUTHORS Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V.,
 Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J.,
 Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M.,
 Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N.,
 Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
 McCann,R., Waterson,R. and Wilson,R.
 Public Soybean EST Project
 Unpublished (1999)
 CONTACT: Shoemaker R/Public Soybean EST Project
 Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 When it has been determined, an EST from the other end of this
 clone is listed in the 'Other ESTs on clone' field. This clone is
 available through: Biogenetic Services, 801 32nd Ave. Brookings, SD
 57006 USA (phone: 800 423 4163; email: info@biogeneticservices.com)
 High quality sequence stop: 201.

TITLE Location/Qualifiers
 JOURNAL 1..202
 COMMENT /organism="Glycine max"
 /mol_type="mRNA"
 /cultivar="Jack"
 /db_xref="taxon:3847"
 /clone="GENOME SYSTEMS CLONE ID: Gm-cl036-5283"
 /tissue_type="somatic embryos cultured on MSD 20"
 /lab_host="DH10B"
 /clone_lib="Gm-cl036"
 /note="Vector: pSPORT1; Site 1: NotI; Site 2: SalI; This
 cDNA library was constructed from mRNA isolated from
 somatic embryos (age ranging from 2 months to 9 months)
 cultured on MSD 20. The library was prepared using the
 Life Technologies pSuperScript cDNA library construction
 kit. Complementary DNA was synthesized from mRNA using a
 poly (dT) sequence with a NotI restriction site. SalI
 linkers adapters were ligated to the blunt-ended cDNA
 fragments followed by NotI digestion. The cDNA fragments
 were directionally cloned into the NotI-SalI restriction
 site of the pSPORT1 vector. The ligated cDNA fragments
 were transformed into E.coli ElectroMax DH10B host cells.
 This library was constructed in the laboratory of Dr. Lila
 Vodkin by Anu Khanna at the University of Illinois at
 Urbana-Champaign. e-mail: l-vodkin@uiuc.edu"

FEATURES
 source
 ORIGIN
 Query Match 26.1%; Score 18; DB 4; Length 202;

Best Local Similarity 100.0%; Pred. No. 23;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 ACTGTAGGAGTTTCAATT 27
 |||||
 Db 53 ACTGTAGGAGTTTCAATT 36

RESULT 6
 CE590287
 LOCUS CE590287 221 bp DNA linear GSS 28-SEP-2003
 DEFINITION tigr-gss-dog-17000366482122 Dog Library Canis familiaris genomic,
 genomic survey sequence.
 ACCESSION CE590287
 VERSION CE590287.1 GI:36907068
 KEYWORDS GSS.
 SOURCE Canis familiaris (dog)
 ORGANISM Canis familiaris

REFERENCE 1 (bases 1 to 221)
 AUTHORS Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
 Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
 Venter,J.C.
 The dog genome: survey sequencing and comparative analysis
 Science 301 (5641), 1898-1903 (2003)

TITLE Location/Qualifiers
 JOURNAL 1..221
 MEDLINE /organism="Canis familiaris"
 PUBMED /mol_type="genomic DNA"
 COMMENT /strain="Standard Poodle"
 /db_xref="taxon:9615"
 /clone_lib="Dog Library"
 /note="Site 1: BstXI; Libraries were prepared from
 peripheral blood"

FEATURES
 source
 ORIGIN
 Query Match 26.1%; Score 18; DB 9; Length 221;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 CTGTAGGAGTTTCAATT 28
 |||||
 Db 76 CTGTAGGAGTTTCAATT 93

RESULT 7
 AW311778
 LOCUS AW311778 297 bp mRNA linear EST 25-APR-2001
 DEFINITION 5655 MARC 2B0V Bos taurus cDNA 5', mRNA sequence.
 ACCESSION AW311778
 VERSION AW311778.1 GI:6727648
 KEYWORDS EST.
 SOURCE Bos taurus (cow)
 ORGANISM Bos taurus

REFERENCE 1 (bases 1 to 297)
 AUTHORS Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
 Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C.,
 Bennett,G.L., Heaton,M.P., Laegreid,W.M., Rohrer,G.A.,
 Chitko-McKown,C.G., Perte,G., Holt,I., Karamycheva,S., Liang,F.,

Quackenbush, J. and Keele, J.W.
 Sequence evaluation of four pooled-tissue normalized bovine cDNA
 libraries and construction of a gene index for cattle
 Genome Res. 11 (4), 626-630 (2001)
 21180013
 11282978
 COMMENT
 Contact: Smith TPL
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390
 Email: smith@email.marc.usda.gov
 Single pass sequencing. Bases called and trimmed with phred
 v0.980904.e. Vector identified by cross_match with the -minscore 20
 and -minmatch 12 options.
 PCR Primers
 FORWARD: AGGAACACGCTATGACCAT
 BACKWARD: GTTTCCTCCAGTCAGCAGC
 Plate: 141 row: M column: 6
 Seq primer: ATTTAGGTGACACTATAG.
 Location/Qualifiers
 1..297
 /organism="Bos taurus"
 /mol_type="mRNA"
 /db_xref="taxon:9913"
 /tissue_type="pooled"
 /lab_host="DH10B"
 /clone_lib="MARC 2BOV"
 /note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;
 Library made from pooled tissue from testis, thymus,
 semitendinosus muscle, longissimus muscle, pancreas,
 adrenal, and endometrium."
 FEATURES
 source
 Query Match 26.1%; Score 18; DB 2; Length 297;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 16 GGAGTTTCAATTTCAGCA 33
 |||||
 Db 41 GGAGTTTCAATTTCAGCA 58
 |||||
 RESULT 8
 AW311777
 LOCUS
 DEFINITION
 5664 MARC 2BOV Bos taurus cDNA 5', mRNA linear EST 25-APR-2001
 ACCESSION
 AW311777
 VERSION
 AW311777.1 GI:6727647
 KEYWORDS
 EST.
 SOURCE
 Bos taurus (cow)
 ORGANISM
 Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 Bovinae; Bos
 REFERENCE
 1 (bases 1 to 315)
 Smith, T.P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T.,
 Casas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C.,
 Bennett, G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A.,
 Chitko-McKown, C.G., Pettea, G., Holt, I., Karamycheva, S., Liang, F.,
 Quackenbush, J. and Keele, J.W.
 Sequence evaluation of four pooled-tissue normalized bovine cDNA
 libraries and construction of a gene index for cattle
 Genome Res. 11 (4), 626-630 (2001)
 21180013
 11282978
 COMMENT
 Contact: Smith TPL
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390
 Email: smith@email.marc.usda.gov
 Single pass sequencing. Bases called and trimmed with phred

v0.980904.e. Vector identified by cross_match with the -minscore 20
 and -minmatch 12 options.
 PCR Primers
 FORWARD: AGGAACACGCTATGACCAT
 BACKWARD: GTTTCCTCCAGTCAGCAGC
 Plate: 141 row: M column: 5
 Seq primer: ATTTAGGTGACACTATAG.
 Location/Qualifiers
 1..315
 /organism="Bos taurus"
 /mol_type="mRNA"
 /db_xref="taxon:9913"
 /tissue_type="pooled"
 /lab_host="DH10B"
 /clone_lib="MARC 2BOV"
 /note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;
 Library made from pooled tissue from testis, thymus,
 semitendinosus muscle, longissimus muscle, pancreas,
 adrenal, and endometrium."
 FEATURES
 source
 Query Match 26.1%; Score 18; DB 2; Length 315;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 16 GGAGTTTCAATTTCAGCA 33
 |||||
 Db 32 GGAGTTTCAATTTCAGCA 49
 |||||
 RESULT 9
 BB847064/c
 LOCUS
 DEFINITION
 BB847064 RIKEN full-length enriched, adult male kidney Mus musculus
 cDNA clone F530005014 5', mRNA sequence.
 ACCESSION
 BB847064
 VERSION
 BB847064.1 GI:17085439
 KEYWORDS
 EST.
 SOURCE
 Mus musculus (house mouse)
 ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE
 1 (bases 1 to 340)
 Akimura, T., Arakawa, T., Carninci, P., Furuno, M., Hanagaki, T.,
 Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Imotani, K.,
 Ishii, Y., Ito, M., Kawai, J., Kojima, Y., Konno, H., Kouda, M.,
 Matsuyama, T., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R.,
 Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sakazume, N.,
 Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T.,
 Sogabe, Y., Suzuki, H., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
 Tanaka, T., Tomaru, A., Toy, T., Watahiki, A., Yasunishi, A.,
 Muramatsu, M. and Hayashizaki, Y.
 RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura, T., et al.
 2001)
 Unpublished (2001)
 Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
 Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)
 wagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,
 Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T.,
 Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A.
 and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format

sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
 Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y. and Hayashizaki, Y.
 Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.
 e mouse tissues.

FEATURES

source
 1. 340
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="F530005014"
 /sex="male"
 /tissue_type="kidney"
 /dev_stage="adult"
 /lab_host="SOLR"
 /clone_lib="RIKEN full-length enriched, adult male kidney"
 /note="Site 1: XhoI; Site 2: SstI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5'
 GAGAGAGAGCGGCGCACTCGAGTTTCTTTTCTTTT 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5'
 GAGAGAGAGAGTCCAGAGCTCAATTAATTAATTAACCCCGCCCC 3'].
 cDNA was cleaved with XhoI and SstI. "

ORIGIN

Query Match 26.1%; Score 18; DB 2; Length 340;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 13 GTAGAGTTTCAATTCA 30
 Db 306 GTAGAGTTTCAATTCA 289

RESULT 10

LOCUS BY306203 340 bp mRNA linear EST 11-DEC-2002
 DEFINITION BY306203 RIKEN full-length enriched, 12.5 days embryo Rathke's pouches Mus musculus cDNA clone K920008111 5', mRNA sequence.
 ACCESSION BY306203
 VERSION BY306203.1 GI:26496540
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus.
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 340)
 Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaide, I., Oshio, N., Saito, R., Suzuki, H., Yananaka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gofjebori, T., Baldairelli, R., Hill, D.P., Bult, C., Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuoka, H., Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V., Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A., Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A., Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W.J., Pertea, G., Pesole, G.,

Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M., Sandelin, A., Schneider, C., Sempile, C.A., Setou, M., Shimada, K., Sultana, R., Takanaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Iehli, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers, J., Birney, E. and Hayashizaki, Y.
 Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)
 2354683
 12468851
 Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Teurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-rsg@gs.riken.jp, URL: <http://genome.gsc.riken.jp/>
 Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission
 Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
 Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
 Tissues were provided by Michelle Brinkmeier and Sally Camper (Dept. Human Genetics University of Michigan Medical School 4301 MSRB 3 1500 W. Medical Center Dr. Ann Arbor, MI 48109-0638 USA) whose assistance we gratefully acknowledge.
 Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.

FEATURES

source
 1. 340
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="CD-1"
 /db_xref="taxon:10090"
 /clone="K920008111"
 /tissue_type="Rathke's pouches"
 /dev_stage="12.5 days embryo"
 /clone_lib="RIKEN full-length enriched, 12.5 days embryo Rathke's pouches"
 ORIGIN
 Query Match 26.1%; Score 18; DB 5; Length 340;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 20 TTTCATTTTCAGCAGGCT 37
 |||||||||||||||||||||

Db 162 TTTCAATTTCAGCAGGT 145

RESULT 11
BY774046/c

LOCUS
DEFINITION BY774046 RIKEN full-length enriched, 17.5 days embryo whole body
Mus musculus cDNA clone L930080D08 5', mRNA sequence.

ACCESSION
VERSION BY774046

KEYWORDS
SOURCE EST.

ORGANISM
Mus musculus (house mouse)

REFERENCE
AUTHORS Carninci, P., Waki, K., Shiraki, T., Konno, H., Shibata, K., Itoh, M., Aizawa, K., Arakawa, T., Ishii, Y., Sasaki, D., Bono, H., Kondo, S., Sugahara, Y., Saito, R., Osato, N., Fukuda, S., Sato, K., Watahiki, A., Hirozane-Kishikawa, T., Nakamura, M., Shibata, Y., Yasunishi, A., Kikuchi, N., Yoshiki, A., Nakagawa, A., Gustincich, S., Beisel, K., Pavan, W., Aidinis, V., Nakagawa, A., Held, W. A., Iwata, H., Kono, T., Nakauchi, H., Lyons, P., Wells, C., Hume, D. A., Fagioli, M., Hensch, T. K., Brinkmeier, M., Camper, S., Hirota, J., Mombaerts, P., Muramatsu, M., Okazaki, Y., Kawai, J., and Hayashizaki, Y.

TITLE Targeting a complex transcriptome: the construction of the mouse full-length cDNA encyclopedia

JOURNAL
MEDLINE Genome Res. 13 (6B), 1273-1289 (2003)

PUBMED 22703353

COMMENT 12819125
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Sushiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site (http://genome.gsc.riken.jp/) for further details.

FEATURES
source
Location/Qualifiers
1..361
/organism="Mus musculus"
/mol_type="mRNA"
/strains="CS7BL/6J"
/db_xref="taxon:10090"
/clone="L930080D08"
/tissue_type="whole body"
/dev_stage="17.5 days embryo"
/clone_lib="RIKEN full-length enriched, 17.5 days embryo whole body"

ORIGIN
Query Match 26.1%; Score 18; DB 6; Length 361;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 GTAGGAGTTCAATTTC 30
|||||

Db 295 GTAGGAGTTCAATTTC 278
|||||

RESULT 12
BY166908/c

LOCUS
DEFINITION BY166908 RIKEN full-length enriched, bone marrow macrophage Mus musculus cDNA clone I830064P15 5', mRNA sequence.

ACCESSION
VERSION BY166908

KEYWORDS
SOURCE EST.

ORGANISM
Mus musculus (house mouse)

REFERENCE
AUTHORS Carninci, P., Waki, K., Shiraki, T., Konno, H., Shibata, K., Itoh, M., Aizawa, K., Arakawa, T., Ishii, Y., Sasaki, D., Bono, H., Kondo, S., Sugahara, Y., Saito, R., Osato, N., Fukuda, S., Sato, K., Watahiki, A., Hirozane-Kishikawa, T., Nakamura, M., Shibata, Y., Yasunishi, A., Kikuchi, N., Yoshiki, A., Nakagawa, A., Gustincich, S., Beisel, K., Pavan, W., Aidinis, V., Nakagawa, A., Held, W. A., Iwata, H., Kono, T., Nakauchi, H., Lyons, P., Wells, C., Hume, D. A., Fagioli, M., Hensch, T. K., Brinkmeier, M., Camper, S., Hirota, J., Mombaerts, P., Muramatsu, M., Okazaki, Y., Kawai, J., and Hayashizaki, Y.

TITLE Targeting a complex transcriptome: the construction of the mouse full-length cDNA encyclopedia

JOURNAL
MEDLINE Genome Res. 13 (6B), 1273-1289 (2003)

PUBMED 22703353

COMMENT 12819125
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Sushiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site (http://genome.gsc.riken.jp/) for further details.

FEATURES
source
Location/Qualifiers
1..361
/organism="Mus musculus"
/mol_type="mRNA"
/strains="CS7BL/6J"
/db_xref="taxon:10090"
/clone="L930080D08"
/tissue_type="whole body"
/dev_stage="17.5 days embryo"
/clone_lib="RIKEN full-length enriched, 17.5 days embryo whole body"

ORIGIN
Query Match 26.1%; Score 18; DB 6; Length 361;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 GTAGGAGTTCAATTTC 30
|||||

Db 295 GTAGGAGTTCAATTTC 278
|||||

BY166908.1 GI:26303554

EST.
Mus musculus (house mouse)

ORGANISM
Mus musculus

REFERENCE
AUTHORS 1 (bases 1 to 385)
Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nakauchi, H., Osato, N., Saito, R., Suzuki, H., Yamanaka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D. P., Bult, C., Hume, D. A., Quackenbush, J., Schriml, L. M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K. W., Blake, J. A., Bradt, D., Brusic, V., Chothia, C., Corbani, L. E., Cousins, S., Dalia, E., Dragani, T. A., Fletcher, C. F., Forrest, A., Frazer, K. S., Gaasterland, T., Gariboldi, M. F., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I. J., Jarvis, E. D., Kanai, A., Kawai, H., Kawasawa, Y., Kedzierski, R. M., King, B. L., Konagaya, A., Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Maglott, D. R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W. J., Pertea, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring, B. Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C. A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tomita, M., Varadar, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L. G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shingawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S., Rogers, J., Birney, E. and Hayashizaki, Y.

TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

JOURNAL
MEDLINE Nature 420, 563-573 (2002)

PUBMED 12466851

COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Sushiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/
Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Tissues were provided by David A. Hume (Depts. of Biochemistry and Microbiology/Parasitology Institute for Molecular Bioscience University of Queensland Brisbane, Q 4072 Australia) whose assistance we gratefully acknowledge.

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

```

FEATURES
  source
    1..385
      /organism="Mus musculus"
      /mol_type="mRNA"
      /strain="CS7BL/6J"
      /db_xref="taxon:10090"
      /clone="I830064P15"
      /tissue_type="bone marrow"
      /cell_type="macrophage"
      /clone_lib="RIKEN full-length enriched, bone marrow
      macrophage"

ORIGIN
  Query Match      26.1%; Score 18; DB 5; Length 385;
  Best Local Similarity 100.0%; Pred. No. 23;
  Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 13 GTAGGAGTTTCAATTTC 30
    |||||
Db 307 GTAGGAGTTTCAATTTC 290

RESULT 13
BY095388/c 392 bp mRNA linear EST 07-DEC-2002
LOCUS
DEFINITION
  BY095388 RIKEN full-length enriched, 13 days embryo whole body Mus
  musculus cDNA clone K630105M02 5', mRNA sequence.

ACCESSION
  BY095388
VERSION
  BY095388.1 GI:26206005
KEYWORDS
  EST.
SOURCE
  Mus musculus (house mouse)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
  1 (bases 1 to 392)
  Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
  Nikaido, I., Oeato, N., Saito, R., Suzuki, H., Yamanaka, I.,
  Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,
  Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D. P., Bult, C.,
  Hume, D. A., Quackenbush, J., Schriml, L. M., Kanapin, A., Matsuda, H.,
  Batalov, S., Beisel, K. W., Blake, J. A., Bradt, D., Brusic, V.,
  Chothia, C., Corbani, L. E., Cousins, S., Dalla, E., Dragani, T. A.,
  Fletcher, C. F., Forrest, A., Frazer, K. S., Gaasterland, T.,
  Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,
  Gustincich, S., Hirokawa, N., Jackson, I. J., Jarvis, E. D., Kanai, A.,
  Kawai, J. H., Kawasawa, Y., Kedzierski, R. M., King, B. L., Kongaya, A.,
  Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Maglott, D. R.,
  Maltas, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,
  Numata, K., Okido, T., Pavan, W. J., Pertea, G., Pesole, G.,
  Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D., Ramsachandran, S.,
  Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring, B. Z., Ringwald, M.,
  Sandelin, A., Schneider, C., Semple, C. A., Setou, M., Shimada, K.,
  Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tomita, M.,
  Verardo, R., Wagner, L., Wahlstedt, C., Wang, Y., Watanabe, Y.,
  Wells, C., Wilming, L. G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I.,
  Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P.,
  Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,
  Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,
  Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y.,
  Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,
  Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S.,
  Rogers, J., Birney, E. and Hayashizaki, Y.
  Analysis of the mouse transcriptome based on functional annotation
  of 60,770 full-length cDNAs
  Nature 420, 563-573 (2002)
  22354683
  12466851
  PUBLISHED

TITLE
  Yoshinide Hayashizaki
  Laboratory for Genome Exploration Research Group, RIKEN Genomic
  Sciences Center (GSC), Yokohama Institute
  The Institute of Physical and Chemical Research (RIKEN)
  COMMENT

```

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@gsc.riken.jp, URL: <http://genome.gsc.riken.jp/>
 Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S.,
 Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H.,
 Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R.,
 Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K.,
 Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and
 Hayashizaki, Y. Direct Submission
 Computational Analysis of Full-Length Mouse cDNAs Compared with
 Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.
 10 (11), 1757-1771 (2000)
 Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for
 further details.

FEATURES

source

1..392

/organism="Mus musculus"

/mol_type="mRNA"

/strain="CS7BL/6J"

/db_xref="taxon:10090"

/clone="K630105M02"

/tissue_type="whole body"

/dev_stage="13 days embryo"

/clone_lib="RIKEN full-length enriched, 13 days embryo

whole body"

Location/Qualifiers

1..392

/organism="Mus musculus"

/mol_type="mRNA"

/strain="CS7BL/6J"

/db_xref="taxon:10090"

/clone="K630105M02"

/tissue_type="whole body"

/dev_stage="13 days embryo"

/clone_lib="RIKEN full-length enriched, 13 days embryo

whole body"

Location/Qualifiers

1..392

/organism="Mus musculus"

/mol_type="mRNA"

/strain="CS7BL/6J"

/db_xref="taxon:10090"

/clone="K630105M02"

/tissue_type="whole body"

/dev_stage="13 days embryo"

/clone_lib="RIKEN full-length enriched, 13 days embryo

whole body"

Location/Qualifiers

1..392

/organism="Mus musculus"

/mol_type="mRNA"

/strain="CS7BL/6J"

/db_xref="taxon:10090"

/clone="K630105M02"

/tissue_type="whole body"

/dev_stage="13 days embryo"

/clone_lib="RIKEN full-length enriched, 13 days embryo

whole body"

Location/Qualifiers

1..392

/organism="Mus musculus"

/mol_type="mRNA"

/strain="CS7BL/6J"

/db_xref="taxon:10090"

/clone="K630105M02"

/tissue_type="whole body"

/dev_stage="13 days embryo"

/clone_lib="RIKEN full-length enriched, 13 days embryo

whole body"

Location/Qualifiers

1..392

/organism="Mus musculus"

/mol_type="mRNA"

/strain="CS7BL/6J"

/db_xref="taxon:10090"

/clone="K630105M02"

/tissue_type="whole body"

/dev_stage="13 days embryo"

/clone_lib="RIKEN full-length enriched, 13 days embryo

whole body"

Location/Qualifiers

1..392

/organism="Mus musculus"

/mol_type="mRNA"

/strain="CS7BL/6J"

/db_xref="taxon:10090"

/clone="K630105M02"

/tissue_type="whole body"

/dev_stage="13 days embryo"

/clone_lib="RIKEN full-length enriched, 13 days embryo

whole body"

Location/Qualifiers

1..392

/organism="Mus musculus"

/mol_type="mRNA"

/strain="CS7BL/6J"

/db_xref="taxon:10090"

/clone="K630105M02"

/tissue_type="whole body"

/dev_stage="13 days embryo"

/clone_lib="RIKEN full-length enriched, 13 days embryo

whole body"

Location/Qualifiers

1..392

/organism="Mus musculus"

/mol_type="mRNA"

/strain="CS7BL/6J"

/db_xref="taxon:10090"

/clone="K630105M02"

/tissue_type="whole body"

/dev_stage="13 days embryo"

/clone_lib="RIKEN full-length enriched, 13 days embryo

whole body"

Location/Qualifiers

1..392

/organism="Mus musculus"

/mol_type="mRNA"

/strain="CS7BL/6J"

/db_xref="taxon:10090"

/clone="K630105M02"

/tissue_type="whole body"

/dev_stage="13 days embryo"

/clone_lib="RIKEN full-length enriched, 13 days embryo

whole body"

Location/Qualifiers

1..392

/organism="Mus musculus"

/mol_type="mRNA"

/strain="CS7BL/6J"

/db_xref="taxon:10090"

/clone="K630105M02"

/tissue_type="whole body"

/dev_stage="13 days embryo"

/clone_lib="RIKEN full-length enriched, 13 days embryo

whole body"

Location/Qualifiers

1..392

/organism="Mus musculus"

/mol_type="mRNA"

/strain="CS7BL/6J"

/db_xref="taxon:10090"

/clone="K630105M02"

/tissue_type="whole body"

/dev_stage="13 days embryo"

/clone_lib="RIKEN full-length enriched, 13 days embryo

whole body"

Location/Qualifiers

1..392

/organism="Mus musculus"

/mol_type="mRNA"

/strain="CS7BL/6J"

/db_xref="taxon:10090"

/clone="K630105M02"

/tissue_type="whole body"

/dev_stage="13 days embryo"

/clone_lib="RIKEN full-length enriched, 13 days embryo

whole body"

Location/Qualifiers

1..392

/organism="Mus musculus"

/mol_type="mRNA"

/strain="CS7BL/6J"

/db_xref="taxon:10090"

/clone="K630105M02"

/tissue_type="whole body"

/dev_stage="13 days embryo"

/clone_lib="RIKEN full-length enriched, 13 days embryo

whole body"

Location/Qualifiers

1..392

/organism="Mus musculus"

/mol_type="mRNA"

/strain="CS7BL/6J"

/db_xref="taxon:10090"

/clone="K630105M02"

/tissue_type="whole body"

/dev_stage="13 days embryo"

/clone_lib="RIKEN full-length enriched, 13 days embryo

whole body"

Location/Qualifiers

1..392

/organism="Mus musculus"

/mol_type="mRNA"

/strain="CS7BL/6J"

/db_xref="taxon:10090"

/clone="K630105M02"

/tissue_type="whole body"

/dev_stage="13 days embryo"

/clone_lib="RIKEN full-length enriched, 13 days embryo

whole body"

Location/Qualifiers

1..392

/organism="Mus musculus"

/mol_type="mRNA"

/strain="CS7BL/6J"

/db_xref="taxon:10090"

/clone="K630105M02"

/tissue_type="whole body"

/dev_stage="13 days embryo"

/clone_lib="RIKEN full-length enriched, 13 days embryo

whole body"

Location/Qualifiers

1..392

/organism="Mus musculus"

/mol_type="mRNA"

/strain="CS7BL/6J"

/db_xref="taxon:10090"

/clone="K630105M02"

/tissue_type="whole body"

/dev_stage="13 days embryo"

/clone_lib="RIKEN full-length enriched, 13 days embryo

whole body"

Location/Qualifiers

1..392

/organism="Mus musculus"

/mol_type="mRNA"

/strain="CS7BL/6J"

/db_xref="taxon:10090"

/clone="K630105M02"

/tissue_type="whole body"

/dev_stage="13 days embryo"

/clone_lib="RIKEN full-length enriched, 13 days embryo

whole body"

Location/Qualifiers

1..392

/organism="Mus musculus"

/mol_type="mRNA"

/strain="CS7BL/6J"

/db_xref="taxon:10090"

/clone="K630105M02"

/tissue_type="whole body"

/dev_stage="13 days embryo"

/clone_lib="RIKEN full-length enriched, 13 days embryo

whole body"

Location/Qualifiers

1..392

/organism="Mus musculus"

/mol_type="mRNA"

/strain="CS7BL/6J"

Kawaji, H., Kawaawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A., Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R., Maitais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W.J., Perte, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M., Sandelin, A., Schneider, C., Semp, C.A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M., Verardo, R., Wagner, L., Wahlstedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers, J., Birney, E. and Hayashizaki, Y.

Analysis of the mouse transcriptome based on functional annotation

of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

22354683

12466851

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/

Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S.,

Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H.,

Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R.,

Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K.,

Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and

Hayashizaki, Y. Direct Submission

Computational Analysis of Full-length Mouse cDNAs Compared with

Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new

genes. Genome Res. 10 (10), 1617-1630 (2000)

RIKEN integrated sequence analysis (RISA) system--384-format

sequencing pipeline with 384 multicapillary sequencer. Genome Res.

10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA

encyclopedia: real-time sequence clustering for construction of a

nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

cDNA library was prepared and sequenced in Mouse Genome

Encyclopedia Project of Genome Exploration Research Group in Riken

Genomic Sciences Center and Genome Science Laboratory in RIKEN.

Division of Experimental Animal Research in Riken contributed to

prepare mouse tissues.

Tissues were provided by David A. Hume (Depts. of Biochemistry

and Microbiology/Parasitology Institute for Molecular Bioscience

University of Queensland Brisbane, Q 4072 Australia) whose

assistance we gratefully acknowledge.

Please visit our web site (http://genome.gsc.riken.go.jp) for

further details.

Location/Qualifiers

1..395

/organism="Mus musculus"

/mol_type="mRNA"

/strain="CS7BL/6J"

/db_xref="taxon:10090"

/clone="I830018N22"

/tissue type="bone marrow"

/cell_type="macrophage"

/clone_lib="RIKEN full-length enriched, bone marrow

macrophage"

macrophage"

ORIGIN

Query Match 26.1%; Score 18; DB 5; Length 395;

Best Local Similarity 100.0%; Pred. No. 23;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 GTAGGAGTTTCAATTTC 30

|||||

Db 310 GTAGGAGTTTCAATTTC 293

|||||

RESULT 15

AA140626

LOCUS

DEFINITION

MEG0147.M1F Schistosoma mansoni, egg stage Schistosoma mansoni

cDNA clone SME0147 similar to 40S ribosomal protein S7, mRNA

sequence.

AA140626

AA140626.1 GI:1702769

EST.

SOURCE

Schistosoma mansoni

Schistosoma mansoni

Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;

Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma.

1 (bases 1 to 408)

Rabelo, E.M.L., Franco, G.R., Azevedo, V., Pena, H.B., Santos, T.M.,

Meira, W.S.F., Rodrigues, N.A., Ortega, J.M. and Pena, S.D.J.

Analysis of cDNA libraries from different developmental stages of

Schistosoma mansoni with the aim of producing Expressed Sequence

Tags

Unpublished (1996)

Contact: Franco G.R. and Pena S.D.J.

Laboratório de Genética-Bioquímica, Departamento de Bioquímica

Imunologia

Instituto de Ciencias Biológicas, Universidade Federal de Minas

Gerais

Avenida Antonio Carlos 6627, Belo Horizonte, MG, Brazil, 31270-010

Tel: (5531) 4415611

Fax: (5531) 4415409

Email: gfranco@mono.icb.ufmg.br

Seq primer: M13 Forward.

Location/Qualifiers

1..408

/organism="Schistosoma mansoni"

/mol_type="mRNA"

/strain="NMRI"

/db_xref="taxon:6183"

/clone="SME0147"

/dev_stage="egg"

/lab_host="DH5alpha"

/clone_lib="Schistosoma mansoni, egg stage"

/note="Vector: pBluescript SK; Site 1: XhoI; Site 2:

EcoRI; mRNA was extracted from eggs and the library was

constructed and excised according to the manufacturer's

instructions (Uni-Zap XR vector, Stratagene)"

ORIGIN

Query Match 26.1%; Score 18; DB 1; Length 408;

Best Local Similarity 100.0%; Pred. No. 23;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 21 TTCAATTTCAGCAGGTA 38

|||||

Db 185 TTCAATTTCAGCAGGTA 202

|||||

Search completed: November 20, 2005, 09:14:27

Job time : 1805 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 2005, 06:45:33 ; Search time 94 Seconds
(without alignments)
1201.097 Million cell updates/sec

Title: US-10-780-250-4
Perfect score: 69
Sequence: 1 gatgtaatactaggagt.....ttactgtaatactactaca 69

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	31.2	45.2	807	US-09-134-000C-1231	Sequence 1231, Ap
C 2	27.6	40.0	174639	US-09-949-016-16509	Sequence 16509, A
C 3	25.6	37.1	767677	US-09-949-016-12147	Sequence 12147, A
C 4	25.6	37.1	767677	US-09-949-016-17361	Sequence 17361, A
C 5	25.4	36.8	51952	US-08-947-823-1	Sequence 1, Appli
C 6	25.2	36.5	65848	US-09-949-016-13285	Sequence 13285, A
C 7	25.2	36.5	235452	US-09-949-016-13675	Sequence 13675, A
C 8	25	36.2	155617	US-09-949-016-16191	Sequence 16191, A
C 9	24.6	35.7	135667	US-09-949-016-15051	Sequence 15051, A
C 10	24.6	35.7	152486	US-09-949-016-12869	Sequence 12869, A
C 11	24.4	35.4	601	US-09-949-016-135426	Sequence 135426, A
C 12	24.4	35.4	131332	US-09-949-016-15535	Sequence 15535, A
C 13	24	34.8	21080	US-09-949-016-15523	Sequence 15523, A
C 14	24	34.8	117080	US-09-949-016-12627	Sequence 12627, A
C 15	23.8	34.5	537	US-09-270-767-831	Sequence 831, App
C 16	23.8	34.5	537	US-09-270-767-16113	Sequence 16113, A
C 17	23.8	34.5	601	US-09-949-016-119330	Sequence 119330, A
C 18	23.8	34.5	1197	US-09-248-796A-1264	Sequence 1264, Ap
C 19	23.8	34.5	36618	US-09-949-016-16935	Sequence 16935, A
C 20	23.8	34.5	4743	US-09-949-016-13796	Sequence 13796, A
C 21	23.4	33.9	261	US-09-513-999C-21561	Sequence 21561, A
C 22	23.4	33.9	601	US-09-949-016-67881	Sequence 67881, A
C 23	23.4	33.9	2995	US-09-949-016-1886	Sequence 1886, Ap
C 24	23.4	33.9	4215	US-09-620-312D-295	Sequence 295, App
C 25	23.4	33.9	6865	US-09-900-038A-3	Sequence 3, Appli
C 26	23.4	33.9	18303	US-09-949-016-13628	Sequence 13628, A
C 27	23.4	33.9	24263	US-09-949-016-12138	Sequence 12138, A

ALIGNMENTS

RESULT 1

US-09-134-000C-1231/c
; Sequence 1231, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134.000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1231
; LENGTH: 807
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
US-09-134-000C-1231

Query Match	45.2%	Score 31.2;	DB 4;	Length 807;
Best Local Similarity	66.2%	Pred. No. 0.033;		
Matches	45;	Conservative	0;	Mismatches 23; Indels 0; Gaps 0;
Qy	2	ATGCTAATCTAGTAGGAGTTTCAATTTTCAGCAGGTTATCAGACGGCTTTACTGTTAATA	61	
Db	475	ATGGCGCTTCTACAGGAATAATATTTTCATCAGACAGATGACAGGCATTTACTGTTAATA	416	
Qy	62	TCACTACA	69	
Db	415	TACATAAA	408	

RESULT 2

US-09-949-016-16509
; Sequence 16509, Application US/099499016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949.016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498

```
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16509
; LENGTH: 174639
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(174639)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16509

Query Match          40.0%; Score 27.6; DB 4; Length 174639;
Best Local Similarity 63.6%; Pred. No. 5.1;
Matches 42; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

Qy 3 TGCTAATCTAGGAGTTTCAATTTCAGCAGGGTATCAGAACGGCTTTACTGGTAATAT 62
Db 89916 TGATCTATTCTAGGACACAGATTAGCAGTGNATAGAGCGTGTTCATGCGCATTT 89975
Qy 63 CACTAC 68
Db 89976 TTATAC 89981

RESULT 3
US-09-949-016-12147
; Sequence 12147, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12147
; LENGTH: 767677
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(767677)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12147

Query Match          37.1%; Score 25.6; DB 4; Length 767677;
Best Local Similarity 62.5%; Pred. No. 49;
Matches 40; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

Qy 6 TAATACTGTAGGAGTTTCAATTTCAGCAGGGTATCAGAACGGCTTTACTGGTAATATC 65
Db 236794 TCATAGTGTGGAACTTCATCTTCAGCAGCCACTAAGAAGGACTTGGCGCTGAGTTT 236853
Qy 66 TACA 69
Db 236854 TGCA 236857

RESULT 4
US-09-949-016-17361
; Sequence 17361, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17361
; LENGTH: 767677
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(767677)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-17361

Query Match          37.1%; Score 25.6; DB 4; Length 767677;
Best Local Similarity 62.5%; Pred. No. 49;
Matches 40; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

Qy 6 TAATACTGTAGGAGTTTCAATTTCAGCAGGGTATCAGAACGGCTTTACTGGTAATATC 65
Db 236794 TCATAGTGTGGAACTTCATCTTCAGCAGCCACTAAGAAGGACTTGGCGCTGAGTTT 236853
Qy 66 TACA 69
Db 236854 TGCA 236857

RESULT 5
US-08-947-823-1
; Sequence 1, Application US/08947823
; Patent No. 6114605
; GENERAL INFORMATION:
; APPLICANT: Williamson, Valerie M.
; APPLICANT: Kaloshian, Isgouhi
; APPLICANT: Yaghoobi, Jafar
; APPLICANT: Bodeau, John
; APPLICANT: Milligan, Stephen
; TITLE OF INVENTION: Procedures and Materials for Conferring
; Pest Resistance in Plants
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/947,823
; FILING DATE: 09-OCT-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US97/18802
; FILING DATE: 09-OCT-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/028,191
; FILING DATE: 10-OCT-1996
; ATTORNEY/AGENT INFORMATION:
```

RESULT 9
 US-09-949-016-15051/c
 ; Sequence 15051, Application US/09949016
 ; Patent No. 6812339
 ; GENERAL INFORMATION:
 ; APPLICANT: VENTER, J. Craig et al.
 ; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
 ; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
 ; FILE REFERENCE: CL001307
 ; CURRENT APPLICATION NUMBER: US/09/949,016
 ; CURRENT FILING DATE: 2000-04-14
 ; PRIOR APPLICATION NUMBER: 60/241,755

; PRIOR FILING DATE: 2000-10-20
 ; PRIOR APPLICATION NUMBER: 60/237,768
 ; PRIOR FILING DATE: 2000-10-03
 ; PRIOR APPLICATION NUMBER: 60/231,498
 ; PRIOR FILING DATE: 2000-09-08
 ; NUMBER OF SEQ ID NOS: 207012
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 15051
 ; LENGTH: 135667
 ; TYPE: DNA
 ; ORGANISM: Human
 US-09-949-016-15051

Query Match 35.7%; Score 24.6; DB 4; Length 135667;
 Best Local Similarity 61.9%; Pred. No. 66;
 Matches 39; Conservative 0; Mismatches 24; Indels 0; Gaps 0;
 Qy 3 TGCTAATCTAGGAGTTTCAATTTACAGCAGGGTATCAGAACGGCTTTACTGGTAATAT 62
 Db 119365 TGCTAAGATTGCAGGAGGTTTCTAATAAGGAGTTTACAAAAGGACTGTTATCTTAAAC 119306
 Qy 63 CAC 65
 Db 119305 AAC 119303

RESULT 10
 US-09-949-016-12869/c
 ; Sequence 12869, Application US/09949016
 ; Patent No. 6812339
 ; GENERAL INFORMATION:
 ; APPLICANT: VENTER, J. Craig et al.
 ; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
 ; FILE REFERENCE: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
 ; FILE REFERENCE: CL001307
 ; CURRENT APPLICATION NUMBER: US/09/949,016
 ; CURRENT FILING DATE: 2000-04-14
 ; PRIOR APPLICATION NUMBER: 60/241,755
 ; PRIOR FILING DATE: 2000-10-20
 ; PRIOR APPLICATION NUMBER: 60/237,768
 ; PRIOR FILING DATE: 2000-10-03
 ; PRIOR APPLICATION NUMBER: 60/231,498
 ; PRIOR FILING DATE: 2000-09-08
 ; NUMBER OF SEQ ID NOS: 207012
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 12869
 ; LENGTH: 152486
 ; TYPE: DNA
 ; ORGANISM: Human
 US-09-949-016-12869

Query Match 35.7%; Score 24.6; DB 4; Length 152486;
 Best Local Similarity 61.9%; Pred. No. 69;
 Matches 39; Conservative 0; Mismatches 24; Indels 0; Gaps 0;
 Qy 3 TGCTAATCTAGGAGTTTCAATTTACAGCAGGGTATCAGAACGGCTTTACTGGTAATAT 62
 Db 88184 TGCTAAGATTGCAGGAGGTTTCTAATAAGGAGTTTACAAAAGGACTGTTATCTTAAAC 88125
 Qy 63 CAC 65
 Db 88124 AAC 88122

RESULT 11
 US-09-949-016-135426/c
 ; Sequence 135426, Application US/09949016
 ; Patent No. 6812339
 ; GENERAL INFORMATION:
 ; APPLICANT: VENTER, J. Craig et al.
 ; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
 ; FILE REFERENCE: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
 ; FILE REFERENCE: CL001307

; CURRENT APPLICATION NUMBER: US/09/949,016
 ; CURRENT FILING DATE: 2000-04-14
 ; PRIOR APPLICATION NUMBER: 60/241,755
 ; PRIOR FILING DATE: 2000-10-20
 ; PRIOR APPLICATION NUMBER: 60/237,768
 ; PRIOR FILING DATE: 2000-10-03
 ; PRIOR APPLICATION NUMBER: 60/231,498
 ; PRIOR FILING DATE: 2000-09-08
 ; NUMBER OF SEQ ID NOS: 207012
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 135426
 ; LENGTH: 601
 ; TYPE: DNA
 ; ORGANISM: Human
 US-09-949-016-135426

Query Match 35.4%; Score 24.4; DB 4; Length 601;
 Best Local Similarity 73.8%; Pred. No. 12;
 Matches 31; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
 Qy 26 TTTCAGCAGGGTATCAGAACGGCTTTACTGGTAATATCACTA 67
 Db 392 TTGAGAAGGGTACCACACGAGCTTTACAGATAGGTCACCA 351

RESULT 12

US-09-949-016-15535/c
 ; Sequence 15535, Application US/09949016
 ; Patent No. 6812339
 ; GENERAL INFORMATION:
 ; APPLICANT: VENTER, J. Craig et al.
 ; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
 ; FILE REFERENCE: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
 ; FILE REFERENCE: CL001307
 ; CURRENT APPLICATION NUMBER: US/09/949,016
 ; CURRENT FILING DATE: 2000-04-14
 ; PRIOR APPLICATION NUMBER: 60/241,755
 ; PRIOR FILING DATE: 2000-10-20
 ; PRIOR APPLICATION NUMBER: 60/237,768
 ; PRIOR FILING DATE: 2000-10-03
 ; PRIOR APPLICATION NUMBER: 60/231,498
 ; PRIOR FILING DATE: 2000-09-08
 ; NUMBER OF SEQ ID NOS: 207012
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 15535
 ; LENGTH: 131332
 ; TYPE: DNA
 ; ORGANISM: Human
 US-09-949-016-15535

Query Match 35.4%; Score 24.4; DB 4; Length 131332;
 Best Local Similarity 73.8%; Pred. No. 78;
 Matches 31; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
 Qy 26 TTTCAGCAGGGTATCAGAACGGCTTTACTGGTAATATCACTA 67
 Db 74152 TTGAGAAGGGTACCACACGAGCTTTACAGATAGGTCACCA 74111

RESULT 13

US-09-949-016-15523
 ; Sequence 15523, Application US/09949016
 ; Patent No. 6812339
 ; GENERAL INFORMATION:
 ; APPLICANT: VENTER, J. Craig et al.
 ; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
 ; FILE REFERENCE: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
 ; FILE REFERENCE: CL001307
 ; CURRENT APPLICATION NUMBER: US/09/949,016
 ; CURRENT FILING DATE: 2000-04-14
 ; PRIOR APPLICATION NUMBER: 60/241,755
 ; PRIOR FILING DATE: 2000-10-20
 ; PRIOR APPLICATION NUMBER: 60/237,768

```
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15523
; LENGTH: 21080
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-15523

Query Match      34.8%; Score 24; DB 4; Length 21080;
Best Local Similarity 64.3%; Pred. No. 59;
Matches 36; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

Qy 14 TAGGAGTTTCAATTTCAGCAGGGTATCAGAACGGCTTTACTGGTAATATCACTACA 69
Db 1171 TATTGTTTCACTTAGTCTTGATTTTGGAGGGTTGTACTAATAATAATAATTACA 1226

RESULT 14
US-09-949-016-12627
; Sequence 12627, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12627
; LENGTH: 117080
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(117080)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12627

Query Match      34.8%; Score 24; DB 4; Length 117080;
Best Local Similarity 64.3%; Pred. No. 1.1e+02;
Matches 36; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

Qy 14 TAGGAGTTTCAATTTCAGCAGGGTATCAGAACGGCTTTACTGGTAATATCACTACA 69
Db 49171 TATTGTTTCACTTAGTCTTGATTTTGGAGGGTTGTACTAATAATAATAATTACA 49226

RESULT 15
US-09-270-767-831/c
; Sequence 831, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 831
; LENGTH: 537
; TYPE: DNA
```

```
; ORGANISM: Drosophila melanogaster
US-09-270-767-831

Query Match      34.5%; Score 23.8; DB 4; Length 537;
Best Local Similarity 59.7%; Pred. No. 20;
Matches 40; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

Qy 2 ATGCTAATACTGTAGGAGTTTCAATTTCAGCAGGGTATCAGAACGGCTTTACTGGTAATA 61
Db 177 AAGGTCAATATTGGATGAGTTTGATGCTCAGGATGGTTTAAAGAAAGCTGTACAATGTCATT 118

Qy 62 TCACTAC 68
Db 117 TCCGTAC 111

Search completed: November 20, 2005, 08:01:18
Job time : 101 secs
```

THIS PAGE BLANK (USPTO)